

Prak-Shastri 1st Semester

Computer Science

Unit: 2nd

Chapter 9

Encoding schemes: Information Interchange (ASCII)

The American Standard Code for Information Interchange (ASCII) is a character encoding scheme that represents text in computers and other electronic devices. It uses a set of 128 characters, including letters, numbers, punctuation marks, and control characters, to represent text.

ASCII was first developed in the 1960s and is still widely used today, particularly in the United States and other parts of the world that use the Latin alphabet. It is the standard encoding scheme for text files in many operating systems and applications.

The ASCII character set consists of 128 characters, including:

- The uppercase and lowercase letters of the English alphabet (A-Z and a-z)
- The digits 0-9
- Punctuation marks and special characters, such as !, @, #, \$, %, &, etc.
- Control characters, such as the newline, tab, and backspace characters, which are used to control the formatting and layout of text.

ASCII is a 7-bit encoding scheme, which means that it uses 7 bits (or 1 byte) to represent each character. This allows for a total of 128 different characters to be represented. ASCII has been widely used for decades and still in use in some systems and protocols. However, it only supports the basic English alphabet and a limited set of special characters, which makes it not suitable for representing more complex characters, such as those used in other languages and scripts. Unicode and UTF-8 are more widely used encoding schemes that can represent a much larger set of characters.

Encoding schemes: Unicode

Unicode is a character encoding standard that represents text in computers and other electronic devices. It is a universal character set that includes almost every character from every writing system in the world, including letters, numbers, punctuation marks, and symbols.

Unicode was first developed in the late 1980s and has since become the standard character encoding scheme for text files in many operating systems and applications. It is designed to be backward-compatible with ASCII and extends the number of characters that can be represented from 128 to over 130,000.

The Unicode character set consists of a vast number of characters, including:

- Characters from the basic Latin alphabet (A-Z and a-z)
- Characters from other scripts such as Greek, Chinese, Japanese, Arabic, etc.
- Mathematical symbols, emoji and various types of punctuation marks
- Control characters, such as the newline, tab, and backspace characters, which are used to control the formatting and layout of text.

Unicode uses a variable-width encoding which means that it uses a different number of bytes to represent different characters. Characters that are used more frequently, such as the basic Latin alphabet, are encoded using fewer bytes, while less common characters, such as those from non-Latin scripts, are encoded using more bytes. Unicode can be encoded using different formats, the most common formats are UTF-8, UTF-16, and UTF-32.

UTF-8 is an encoding of Unicode that uses 8-bit sequences to represent characters, it is backward-compatible with ASCII and it's the most popular encoding format on the internet. UTF-16 uses 16-bit sequences to represent characters, it's mostly used in systems that need to handle a large number of characters, such as Windows. UTF-32 uses 32-bit sequences to represent characters, it's used in systems that need to handle a large number of characters and it's more efficient than UTF-16 in terms of memory usage.

Encoding schemes: Indian Script Code for Information Interchange (ISCII)

Indian Script Code for Information Interchange (ISCII) is a character encoding standard that is specifically designed to represent the scripts used in

India, such as Devanagari, Bengali, Gujarati, Gurmukhi, Kannada, Malayalam, Oriya, Tamil, Telugu, and Urdu. It was developed in the early 1990s by the Indian government's National Informatics Centre (NIC) and is still used in some systems and applications today.

ISCII is based on the 8-bit ASCII character set, which means that it uses 8 bits (or 1 byte) to represent each character. It includes a set of characters for each of the Indian scripts, as well as additional characters for punctuation marks, diacritics, and other special characters.

The ISCII character set consists of several different code pages, each of which is designed for a specific script. The code pages include:

- Devanagari script code page
- Bengali script code page
- Gujarati script code page
- Gurmukhi script code page
- Kannada script code page
- Malayalam script code page
- Oriya script code page
- Tamil script code page
- Telugu script code page
- Urdu script code page

ISCII has been widely used in India for decades, but it has some limitations, as it can't handle complex scripts and it's not compatible with other encoding standards, like Unicode. Unicode is now more widely used to represent Indian scripts, as it can handle a much larger set of characters and it's more widely supported. However, some legacy systems still use ISCII encoding and many Indian Language Processing Applications still use ISCII for backward compatibility.

Practice Questions:

1. What is the purpose of ASCII encoding?

2. How many bits are used to represent each character in ASCII?
3. What is the range of characters that can be represented in ASCII?
4. How does ASCII differ from other encoding schemes?
5. What is the origin of the ASCII encoding standard?
6. What is the relationship between ASCII and the American Standard Code for Information Interchange?
7. How is the ASCII code table organized?
8. What is the difference between ASCII and Unicode?
9. What is the ASCII code for the letter 'A'?
10. What is the ASCII code for the digit '0'?
11. How are special characters represented in ASCII?
12. Can ASCII encode non-English characters?
13. How is the ASCII encoding used in modern computing?
14. How is ASCII used in networking protocols?
15. What is the ASCII code for the space character?
16. Can ASCII be extended to include more characters?
17. What is the ASCII code for the newline character?
18. How does ASCII handle diacritical marks?
19. What is the ASCII code for the character '\$'?
20. How does ASCII handle upper and lower case letters?
21. What is the ASCII code for the character '@'?
22. Are ASCII codes standardized across different platforms?
23. What is the ASCII code for the character '#'?
24. How does ASCII handle non-printable characters?
25. What is the ASCII code for the character '&'?