Unicode in Computer Network

Last Updated: 12-08-2019

Unicode is a universal encoding system to provide a comprehensive character set and was created by the Unicode consortium (a group of multilingual software manufacturers). Unicode simplifies software localization and improves multilingual text processing. It overcomes the difficulty inherent in ASCII and extended ASCII.

Unicode has standardizes script behavior which allows any combination of characters, drawn from any combination of scripts and languages, to co-exist in a single document. Unicode defines multiple encodings of its single character set: UTF-7, UTF-8, UTF-16, and UTF-32. **Conversion** of data among these encodings is **lossless**.

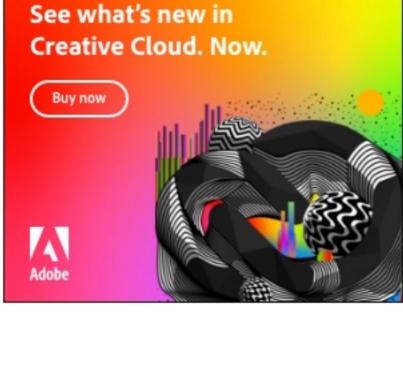Unicode was originally a 2-byte character set. Unicode version 3, however is a **4-byte** code and is fully compatible with ASCII and extended ASCII.

These all support encoding the same set of characters.

- **UTF-8** uses anywhere from 1 to 4 bytes per character depending on character, but ASCII take only 1 byte and 4 bytes for unusual ones.
- **UTF-16** uses 2 bytes for most characters, while very unusual characters take 4.
- **UTF-32** uses 4 bytes per character. We can calculate the number of characters in a UTF-32 string by only counting bytes.

The notation uses hexadecimal digits in format as follows.

U-XXXXXXXX –

The numbering goes from **U-00000000 to U-FFFFFFFF**. Unicode divides the available space codes into planes. A plane is a continuous group of 65,536 code points. The most significant 16 bits define the plane (i.e. number of planes = 65,535) and each plane can define up to 65,536 character or symbols.

Types of Plane –

1. **Basic multilingual plane (BMP) –** Plane 0000, the basic multilingual plane is designed to be compatible with the previous 16-bit Unicode. The most significant 16-bits in this plane are all zeroes. It mostly defines character sets in different languages with the exception of some control and special characters. It is represented by U+XXXX where XXXX are least significant 16-bits, eig, : U+0900 to U+09FF reserved for Devanagari, Bengali U+2200 to U+22FF reserved for mathematical operation etc.

2. **Supplementary multilingual plane (SMP) –** Plane 0001, the supplementary multilingual plane, is designed to provide more codes for those multilingual characters that are excluded in the BMP. Example: 10140-1018F are reserved for Ancient Greek Numbers.

3. **Supplementary ideography plane (SIP) –** Plane 0002, the supplementary ideography plane, is designed to provide codes for ideographic symbols, symbols that provide an idea in contrast to a sound, e.g., 20000-2A6DF are reserved for CJK Unified Extension B

4. **Supplementary special plane (SSP) –** Plane 000E, the supplementary special plane, is used for special characters, e.g., E0000-E007F are reserved for tags.

5. **Private use planes (PUPs) –** Planes 000F and 0010, private use planes are for private use. They are used by fonts internally to refer to auxiliary glyphs.

Reference –

Unicode – msdn.microsoft

Data Communication and Networking – Forouzan

This article is contributed by **Himanshi**. If you like GeeksforGeeks and would like to contribute, you can also write an article using contribute.geeksforgeeks.org or mail your article to contribute@geeksforgeeks.org. See your article appearing on the GeeksforGeeks main page and help other Geeks.

Please write comments if you find anything incorrect, or you want to share more information about the topic discussed above.

Attention reader! Don't stop learning now. Get hold of all the important CS Theory concepts for SDE interviews with the **CS Theory Course** at a student-friendly price and become industry ready.

Recommended Posts:

Basic Network Attacks in Computer Network
Difference between Storage Area Network (SAN) and Network Attached Storage (NAS)
Differences between Wireless Adhoc Network and Wireless Sensor Network
Difference between Next Generation Network and Traditional Network
Difference between Software Defined Network and Traditional Network
Computer Network | Leaky bucket algorithm
HMAC Algorithm in Computer Network
Introduction of Firewall in Computer Network
Circuit Switching in Computer Network
Packet Switching and Delays in Computer Network
Routing v/s Routed Protocols in Computer Network
Difference between Unicast, Broadcast and Multicast in Computer Network
Introduction of MAC Address in Computer Network
Types of Server Virtualization in Computer Network
Hamming Code in Computer Network
MAC Filtering in Computer Network
DHCP Relay Agent in Computer Network
Controlled Access Protocols in Computer Network
Redundant Link problems in Computer Network
Multiple Access Protocols in Computer Network

Article Tags :   Computer Networks   CBSE - Class 11   school-programming

Practice Tags :   Computer Networks

👍 Be the First to upvote.

0

No votes yet.

☐ To-do   ☐ Done

Improve Article

Please write to us at contribute@geeksforgeeks.org to report any issue with the above content.

Previous	Next
⏮ TCP Timers	Collision Avoidance in wireless networks ⏭

Writing code in comment? Please use ide.geeksforgeeks.org, generate link and share the link here.

GeeksforGeeks

5th Floor, A-118, Sector-136, Noida, Uttar Pradesh - 201305
feedback@geeksforgeeks.org

Company
About Us
Careers
Privacy Policy
Contact Us

Learn
Algorithms
Data Structures
Languages
CS Subjects
Video Tutorials

Practice
Courses
Company-wise
Topic-wise
How to begin?

Contribute
Write an Article
Write Interview Experience
Internships
Videos