**Topic: Character sets – ASCII code and Unicode**

Reading Time: 20 mins

**·        Note\* Highlight important/core points while reading**

·        Read the content and write the answers given in the document in your words, to get the solid grip on topic.

**Character Sets – ASCII Code and Unicode**

Computers store text as **binary numbers**. However, to represent letters, numbers, and symbols, a **character encoding system** is needed. Two commonly used character sets are **ASCII (American Standard Code for Information Interchange) and Unicode**.

**1. ASCII Code**

ASCII (pronounced "ask-ee") is one of the earliest character encoding standards. It assigns **a unique binary number to each character** in the English alphabet, digits, punctuation marks, and control characters.

**Part of the ASCII Code Table (7-bit ASCII)**

|  |  |  |  |
| --- | --- | --- | --- |
| **Character** | **Decimal (Denary)** | **Binary** | **Hexadecimal** |
| A | 65 | 1000001 | 41 |
| B | 66 | 1000010 | 42 |
| C | 67 | 1000011 | 43 |
| a | 97 | 1100001 | 61 |
| b | 98 | 1100010 | 62 |
| c | 99 | 1100011 | 63 |
| 0 | 48 | 0110000 | 30 |
| 1 | 49 | 0110001 | 31 |
| Space | 32 | 0100000 | 20 |

**Features of ASCII**

* Uses **7 bits per character**, allowing **128 different characters (0 to 127 in decimal)**.

Includes uppercase and lowercase English letters, digits (0-9), punctuation, and control codes (e.g., Enter, Backspace).

**2. Extended ASCII (8-bit ASCII)**

To support **non-English characters and graphical symbols**, **Extended ASCII** was developed.

* Uses **8 bits per character**, allowing **256 characters (0 to 255 in decimal, 0 to FF in hexadecimal)**.
* Includes additional symbols, accents (e.g., é, ñ, ö), and box-drawing characters.

**Example Extended ASCII Table (8-bit)**

|  |  |  |  |
| --- | --- | --- | --- |
| **Character** | **Decimal** | **Binary** | **Hexadecimal** |
| Ç | 199 | 11000111 | C7 |
| é | 233 | 11101001 | E9 |
| ñ | 241 | 11110001 | F1 |
| ß | 223 | 11011111 | DF |
| © | 169 | 10101001 | A9 |

**3. Unicode**

ASCII is **limited** because it cannot represent all world languages. Unicode was created as a **universal character encoding system**.

**Features of Unicode:**

·         Uses **16-bit, 32-bit, or variable-length encoding**.

·         Supports over **143,000 characters** from **different languages** and **symbols**.

·         Used in **modern computing (web, programming, databases, etc.)**.

**Sample Unicode Characters**

|  |  |  |
| --- | --- | --- |
| **Character** | **Unicode Code Point** | **Description** |
| A | U+0041 | Latin Capital Letter A |
| Ω | U+03A9 | Greek Capital Letter Omega |
| 文 | U+6587 | Chinese Character for "Text" |
| 💻 | U+1F4BB | Laptop Emoji |
| ☀ | U+2600 | Sun Symbol |

**Comparison: ASCII vs. Unicode**

|  |  |  |
| --- | --- | --- |
| **Feature** | **ASCII** | **Unicode** |
| Number of Bits | 7-bit (Standard ASCII), 8-bit (Extended ASCII) | 16-bit, 32-bit, or variable length |
| Characters Supported | 128 (7-bit), 256 (8-bit) | 143,000+ characters |
| Supports Non-English? | No (Standard ASCII), Yes (Extended ASCII) | Yes (All languages) |
| Emoji Support | No | Yes |
| Used in Modern Systems? | Limited | Yes (Web, Programming, Databases) |

### Write your Answers on your Notebook and Verify it on Next Screen

**Q6: Convert the ASCII character 'b' to decimal, binary, and hexadecimal.**

**Q7: How does Unicode represent characters differently from ASCII?**

**Q8: What is the Unicode code point for the 'Ω' (Greek Capital Letter Omega) symbol?**

**Q9: How many bits are required to store a standard ASCII character and an extended ASCII character?**

**Q10: Why is ASCII not suitable for storing text in multiple languages?**

**6. Answer:**

* **ASCII Code of 'b'** = **98** (decimal)
* **Binary Representation** = **1100010₂**
* **Hexadecimal Representation** = **62₁₆**

**Answer:** 'b' → **98 (decimal), 1100010₂ (binary), 62₁₆ (hexadecimal).**

**7. Answer:**

* ASCII uses **7-bit (Standard)** or **8-bit (Extended ASCII)** encoding, limiting it to **128 or 256 characters**.
* Unicode uses **16-bit, 32-bit, or variable-length encoding**, allowing it to support **over 143,000 characters**.
* Unicode includes **all world languages, symbols, and emojis**, making it more versatile for global communication.

**Answer:** Unicode supports **more characters** and **different languages** using **larger encoding sizes** compared to ASCII.

**8. Answer:**

* The Unicode code point for **'Ω' (Greek Capital Letter Omega)** is **U+03A9**.

**Answer:** **Ω → U+03A9** in Unicode.

**9. Answer:**

* **Standard ASCII** requires **7 bits** per character.
* **Extended ASCII** requires **8 bits** per character.

**Answer:**

* **7 bits for Standard ASCII**
* **8 bits for Extended ASCII**

**10. Answer:**

* ASCII supports **only 128 (Standard) or 256 (Extended) characters**, which **limits** it to English and a few other symbols.
* It **cannot represent** characters from other languages like **Chinese, Arabic, Hindi, or Cyrillic scripts**.
* Unicode, however, supports **thousands of characters** and is widely used for multilingual text storage.

**Answer:** ASCII **lacks support for multiple languages**, making it **unsuitable for global text representation**.