

DEFINITION

ASCII (American Standard Code What is ASCII?



ASCII (American Standard Code for Information Interchange) is the most common character encoding format for text data in computers and on the internet. In standard ASCII-encoded data, there are unique values for 128 alphabetic, numeric or special additional characters and control codes.

ASCII encoding is based on character encoding used for telegraph data. The [American National Standards Institute](#) first published it as a standard for computing in 1963.

Characters in ASCII encoding include upper- and lowercase letters A through Z, numerals 0 through 9 and basic punctuation symbols. It also uses some non-printing control characters that were originally intended for use with teletype printing terminals.

ASCII characters may be represented in the following ways:

- as pairs of [hexadecimal](#) digits -- base-16 numbers, represented as 0 through 9 and A through F for the decimal values of 10-15;
- as three-digit octal (base 8) numbers;
- as decimal numbers from 0 to 127; or
- as 7-bit or 8-bit [binary](#)

| Character | Hexadecimal | Octal | Decimal | Binary (7 bit) | Binary (8 bit) |
|-----------|-------------|-------|---------|----------------|----------------|
| m | 0x6D | /155 | 109 | 110 1101 | 0110 1101 |

ASCII characters were initially encoded into 7 bits and stored as 8-bit characters with the [most significant bit](#) -- usually, the left-most bit -- set to 0.

Why is ASCII important?

ASCII was the first major character encoding standard for data processing. Most modern computer systems use [Unicode](#), also known as the Unicode Worldwide Character Standard. It's a character encoding standard that includes ASCII encodings.

The Internet Engineering Task Force ([IETF](#)) adopted ASCII as a standard for internet data when it published "ASCII format for Network Interchange" as [RFC 20](#) in 1969. That request for comments ([RFC](#)) document standardized the use of ASCII for internet data and was accepted as a full standard in 2015.

ASCII encoding is technically obsolete, having been replaced by Unicode. Yet, ASCII characters use the same encoding as the first 128 characters of the Unicode Transformation Format 8, so ASCII text is compatible with UTF-8.

In 2003, the IETF standardized the use of UTF-8 encoding for all web content in [RFC 3629](#).

Almost all computers now use ASCII or Unicode encoding. The exceptions are some IBM mainframes that use the proprietary 8-bit code called Extended Binary Coded Decimal Interchange Code ([EBCDIC](#)).

How does ASCII work?

ASCII offers a universally accepted and understood character set for basic data communications. It enables developers to design interfaces that both humans and computers understand. ASCII codes a string of data as ASCII characters that can be interpreted and displayed as readable plain text for people and as data for computers.

Programmers use the design of the ASCII character set to simplify certain tasks. For example, using ASCII character codes, changing a single bit easily converts text from uppercase to lowercase.

The capital letter "A" is represented by the binary value:

```
0100 0001
```

The lowercase letter "a" is represented by the binary value:

```
0110 0001
```

| Character | Binary | Decimal | Hexadecimal |
|-----------|-----------|---------|-------------|
| A | 0100 0001 | 65 | 0x41 |
| a | 0110 0001 | 97 | 0x61 |

The difference is the third most significant bit. In [decimal and hexadecimal](#), this corresponds to:

The difference between upper- and lowercase characters is always 32 (0x20 in hexadecimal), so converting from upper- to lowercase and back is a matter of adding or subtracting 32 from the ASCII character code.

ASCII advantages and disadvantages

After more than half a century of use, the advantages and disadvantages of using ASCII character encoding are well understood. That is one of the encoding format's great strengths.

Advantages

- **Universally accepted.** ASCII character encoding is universally understood. Except for the IBM mainframes that use EBCDIC encoding, it is universally implemented in computing through the Unicode standard. Unicode character encoding replaces ASCII encoding, but it is backward-compatible with ASCII.
- **Compact character encoding.** Standard codes can be expressed in 7 bits. This means data that can be expressed in the standard ASCII character set requires only as many bytes to store or send as the number of characters in the data.
- **Efficient for programming.** The character codes for letters and numbers are well adapted to programming techniques for manipulating text and using numbers for calculations or storage as raw data.

Disadvantages

- **Limited character set.** Even with extended ASCII, only 255 distinct characters can be represented. The characters in a standard character set are enough for English language communications. But even with the diacritical marks and Greek letters supported in extended ASCII, it is difficult to accommodate languages that do not use the Latin alphabet.

- **Inefficient character encoding.** Standard ASCII encoding is efficient for English language and numerical data. Representing characters from other alphabets requires more overhead such as escape codes.

