

# Text file

---

A **text file** (sometimes spelled "textfile": an old alternate name is "flatfile") is a kind of computer file that is structured as a sequence of lines. A text file exists within a computer file system. The end of a text file is often denoted by placing one or more special characters, known as an end-of-file marker, after the last line in a text file.

"Text file" refers to a type of container, while plain text refers to a type of content. Text files can contain plain text, but they are not limited to such.

At a generic level of description, there are two kinds of computer files: text files and binary files.<sup>[1]</sup>

## Data storage

Because of their simplicity, text files are commonly used for storage of information. They avoid some of the problems encountered with other file formats, such as endianness, padding bytes, or differences in the number of bytes in a machine word. Further, when data corruption occurs in a text file, it is often easier to recover and continue processing the remaining contents. A disadvantage of text files is that they usually have a low entropy, meaning that the information occupies more storage than is strictly necessary.

A simple text file needs no additional metadata to assist the reader in interpretation, and therefore may contain no data at all, which is a case of zero byte file.

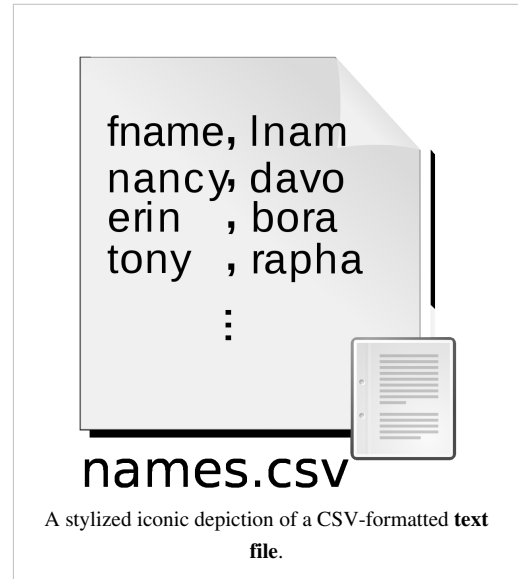
## Formats

### ASCII

The ASCII standard allows ASCII-only text files (unlike most other file types) to be freely interchanged and readable on Unix, Macintosh, Microsoft Windows, DOS, and other systems. These differ in their preferred line ending convention and their interpretation of values outside the ASCII range (their character encoding).

### MIME

Text files usually have the MIME type "text/plain", usually with additional information indicating an encoding. Prior to the advent of Mac OS X, the Mac OS system regarded the content of a file (the data fork) to be a text file when its resource fork indicated that the type of the file was "TEXT". Under the Microsoft Windows operating system, a file is regarded as a text file if the suffix of the name of the file (the "extension") is ".txt". However, many other suffixes are used for text files with specific purposes. For example, source code for computer programs is usually kept in text files that have file name suffixes indicating the programming language in which the source is written.



## .TXT

**.txt** is a file format for files consisting of text usually containing very little formatting (ex: no **bolding** or *italics*). The precise definition of the .txt format is not specified, but typically matches the format accepted by the system terminal or simple text editor. Files with the .txt extension can easily be read or opened by any program that reads text and, for that reason, are considered universal (or platform independent).

The ASCII character set is the most common format for English-language text files, and is generally assumed to be the default file format in many situations. For accented and other non-ASCII characters, it is necessary to choose a character encoding. In many systems, this is chosen on the basis of the default locale setting on the computer it is read on. Common character encodings include ISO 8859-1 for many European languages.

Because many encodings have only a limited repertoire of characters, they are often only usable to represent text in a limited subset of human languages. Unicode is an attempt to create a common standard for representing all known languages, and most known character sets are subsets of the very large Unicode character set. Although there are multiple character encodings available for Unicode, the most common is UTF-8, which has the advantage of being backwards-compatible with ASCII: that is, every ASCII text file is also a UTF-8 text file with identical meaning.

## Standard Windows .txt files

MS-DOS and Windows use a common text file format, with each line of text separated by a two character combination: CR and LF, which have ASCII codes 13 and 10. It is common for the last line of text **not** to be terminated with a CR-LF marker, and many text editors (including Notepad) do not automatically insert one on the last line.

Most Windows text files use a form of ANSI, OEM or Unicode encoding. What Windows terminology calls "ANSI encodings" are usually single-byte ISO-8859 encodings, except for in locales such as Chinese, Japanese and Korean that require double-byte character sets. ANSI encodings were traditionally used as default system locales within Windows, before the transition to Unicode. By contrast, OEM encodings, also known as MS-DOS code pages, were defined by IBM for use in the original IBM PC text mode display system. They typically include graphical and line-drawing characters common in full-screen MS-DOS applications. Newer Windows text files may use a Unicode encoding such as UTF-16LE or UTF-8.

## Rendering

When opened by a text editor, human-readable content is presented to the user. This often consists of the file's plain text visible to the user. Depending on the application, control codes may be rendered either as literal instructions acted upon by the editor, or as visible escape characters that can be edited as plain text. Though there may be plain text in a text file, control characters within the file (especially the end-of-file character) can render the plain text unseen by a particular method.

## Notes and references

[1] Lewis, John (2006). *Computer Science Illuminated*. Jones and Bartlett. ISBN 0763741493.

## External links

- C2: the Power of Plain Text (<http://c2.com/cgi/wiki?PowerOfPlainText>)

# Article Sources and Contributors

**Text file** *Source:* <http://en.wikipedia.org/w/index.php?oldid=440852736> *Contributors:* 16@r, 1c33y37, 4twenty42o, Abdull, Ahoerstemeier, Alex Peppe, Ali, Angela, Ann Stouter, Arctic Fox, Arvindn, Atomice, Bbq man, BirgitteSB, Boffy b, Brandon, CAPS LOCK, CanisRufus, Christopherlin, Colonies Chris, Courcelles, DaGizza, DavidCary, Deb, Delirium, Derek farn, Dim131, Dreftymac, Dysprosia, Ejg930, Evercat, Filzstift, FisherQueen, Franl, Func, Furrykef, Garas, Gerry Ashton, Guyalsfere, Gwern, Harryboyles, Hdante, Head, Incnis Mrsi, Iranway, Iridiumcao, Jc3s5h, Jkl, JoanneB, Karl-Henner, Kieff, Kirun, Kowloonese, Leftovermelody, Leujohn, LittleDan, Lobogris, Loggie, Lovibond, Lysdexia, Maniac18, Markaci, Martarius, Martinkunev, Materialscientist, Matnerd, Mendel, Mglovesfun, Michael Hardy, Mikeblas, Mkilly, Nahum Reduta, Neelix, Newsaholic, OlEnglish, Oli Filth, Patrick, Pengo, Peruvianllama, Philip Trueman, Piet Delpport, Pit, Pulseczar, R. S. Shaw, Radagast83, Rd232, RedWolf, Rodeosmurf, Rpyle731, Rursus, RussBlau, S, Sampl, Shadowjams, Sinar, Sjlegg, TakuyaMurata, Tanár, Tardis, Techmdrn, The Anome, Tide rolls, Unyoyega, Voidvector, Wapcaplet, Wayward, Ww, Yonkie, Ziounclesi, 128 anonymous edits

# Image Sources, Licenses and Contributors

**Image:CsvDelimited001.svg** *Source:* <http://en.wikipedia.org/w/index.php?title=File:CsvDelimited001.svg> *License:* Creative Commons Attribution-Sharealike 2.5 *Contributors:* User:Dreftymac

# License

Creative Commons Attribution-Share Alike 3.0 Unported  
<http://creativecommons.org/licenses/by-sa/3.0/>