

Files



What is a File?




The File Abstraction

Files are a common digital object. A file is a container for a string of bits, with a name and a size. The file itself is an abstraction of a set of fixed-size records stored in disk memory as a tree structure rooted at an index table. Users of files do not see the records, the tree, and the index table.

Peter J. Denning, "The Profession of IT: Systems Abstraction," *CACM*, April 2022, volume 85, number 4, page 24.

Files

- A computer organizes all its data in blocks called files.
- All the data is stored in zeros and ones, because ultimately that is the only possibility that hardware is capable of.
- Everything stored, music, books, movies, pictures, Java programs, computer instructions, everything is a binary file. See examples: [Digital](#) .
- Text files are special binary files in which the bits are understood to be characters. Often 8 bits (aka one byte) is used for a letter, digit, or punctuation.

Files

```
8b 08 40 92 24 ff e7 ee b9
09 41 c4 d7 d1 cc 02 a3 38
51 4f a1 45 19 d3 96 8e cf
30 e3 a5 48 7d 36 65 bc bf
a0 22 4a c1 3a 1f d8 4d 4a
6d 36 1e 83 0f 22 fa c3 4a
50 15 9e 03 4a 4a 7f 03
```

A *computer file* is a block of information stored and maintained by the operating system on a computer. The file remains available for programs to use after the current program has finished (by being stored in secondary storage—the hard drive). Computer files are the modern counterpart of paper documents kept in offices and libraries.

Files

```
8b 08 40 92 24 ff e7 ee b9
09 41 c4 d7 d1 cc 02 a3 38
51 4f a1 45 19 d3 96 8e cf
30 e3 a5 48 7d 36 65 bc bf
a0 22 4a c1 3a 1f d8 4d 4a
6d 36 1e 83 0f 22 fa c3 4a
50 15 9e 03 4a 4a 7f 03
```

A computer file is a one-dimensional, sequence of data. All data is stored by a computer as binary bits. The smallest manageable unit of data is generally the 8-bit byte, so files and other basic structures of data are build from integral numbers of bytes.

(Boolean data—representable by just one-bit—is out of luck; it is not an important enough case to justify the extra overhead.)

Text Files

Text files are a special case of (binary) files. A character encoding is used to interpret the bit patterns of the file as characters of natural language text. Since these files are meant to be read by people, they have lines.

In the past, certain bit patterns were meant to control transmission or display, but except for the line terminator these bit patterns are obsolete or a nuisance, so it is best to avoid them.

```
A B C D E n1 F G n1
```

Do not assume that characters are the same size as bytes.

File System

OS maintains a (hierarchically) organized collection of files. Unix influence: tree-oriented, directory are files, etc.

`/root/directory/subdirectory/file`



File System

