Theory: Files

🕐 7 minutes   7 / 7 problems solved

Start practicing

§1. What is a file

Computers need to store a tremendous amount of data. On most operating systems, this data is stored in **files**.

A **computer file** is a collection of data stored on a disk/another storage device that has a name and can be manipulated by a user or program as a single piece of information. One doesn't have to go far for a good example: this web page is a file; the image you see below is a file; even your favorite YouTube video is a file – no matter which one it is! In fact, any sort of data can be written in a computer file.

A user or a program can open, read, modify, save, and close files. Also, files can be transferred through the particular device they are stored in or through the Internet.

§2. Contents and icons

All modern operating systems store data in files, so even though the way data is organized differs between operating systems, conceptually they are all the same. Let's look at a specific example – this picture stored in a single file:
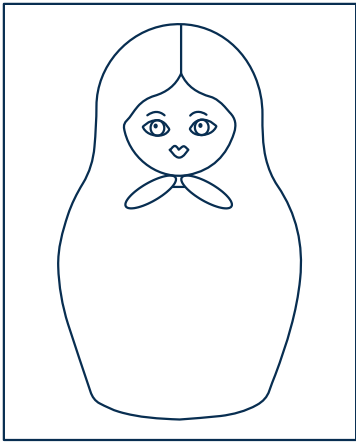
File name ⟶ img_1.jpg ⟵ File extension

File format ⟶ JPEG image - 22KB ⟵ File size

In the graphical user interface of an OS, files are usually represented by **icons**: small pictures that indicate their contents. For example, below are the macOS icons for an image, a Python program, a book and a text file.

Files.jpg   generate_words.py   Mark_Lutz_-_Learning_...5th_Ed.pdf   instruction.txt
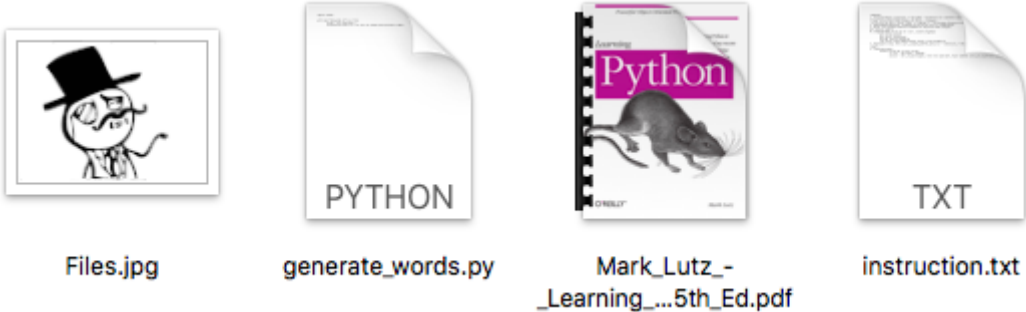
§3. Metadata

Besides the contents, each file has metadata. **Metadata** is the additional information about the file stored in computer memory (you know how important it is for a computer to manage files properly). Let's take a look at some examples of metadata that file can have.

- **Name**: a unique identifier of the file. In our case, it's *img_1.jpg*.

- **Extension**: a filename suffix starting with a dot and indicating the format. For example, *.txt* extension is used for plain text files, *.img* or *.jpg* are used for images and so on.

Current topic:

✓ Files ⋯

Topic is required for:

Introduction to databases ⋯

✓ Files in Python ⋯

- **Format**, or **type**. The file format is indicated by an extension: it tells us what kind of information is stored there. It can be a text message, an image, a video, an audio file, a computer program – you name it. For example, the *.jpg* extension indicates that this file is of the *JPEG image* format. All files on all modern types of OS are organized as 1-D arrays of bytes. The format specifies the rules of how the computer must read and interpret these bytes. For instance, in plain text format, characters are encoded as sequences of bytes, while the bytes of audiofiles, videofiles, and images are interpreted differently.

- **Size**: how much space the file occupies on a device (usually measured in bytes).

§4. Attributes

Let's discuss a special kind of metadata called **file attributes**. Each file has some set of attributes, though this set varies greatly from one OS to another. Each attribute can be either *set* (toggled on) or *cleared* (toggled off). For example, let's look at the following three attributes: these basic attributes are included in *every* version of Windows:

- Hidden (H): when set, makes the file hidden, which means that the user cannot see it by default.

- System (S): when set, indicates that the file is crucial to the system and the computer to operate properly.

- Read-only (R): when set, makes the file read-only, which means users cannot write into it.

Essentially, attributes are data that define the behavior of the file system.

§5. Directories

Files can be organized into catalogs known as **directories,** or **folders**. A directory is a container that can store files and other directories – like an actual office file folder that contains documents and other folders. Usually, directory icons even resemble office file folders – for example, look at these lovely default directory icons for Windows, Linux (Ubuntu) and Mac OS respectively:

A directory inside another directory is called a **subdirectory**. Sometimes, terms **parent** directory and **child** directory are used.

Files and directories inside an operating system are organized hierarchically. A top-most directory with no parents is called **root directory**. It contains many subdirectories that in turn contain more subdirectories, and the story goes on... This process may remind one of a tree, so that's why the top-most directory is called a 'root directory'.
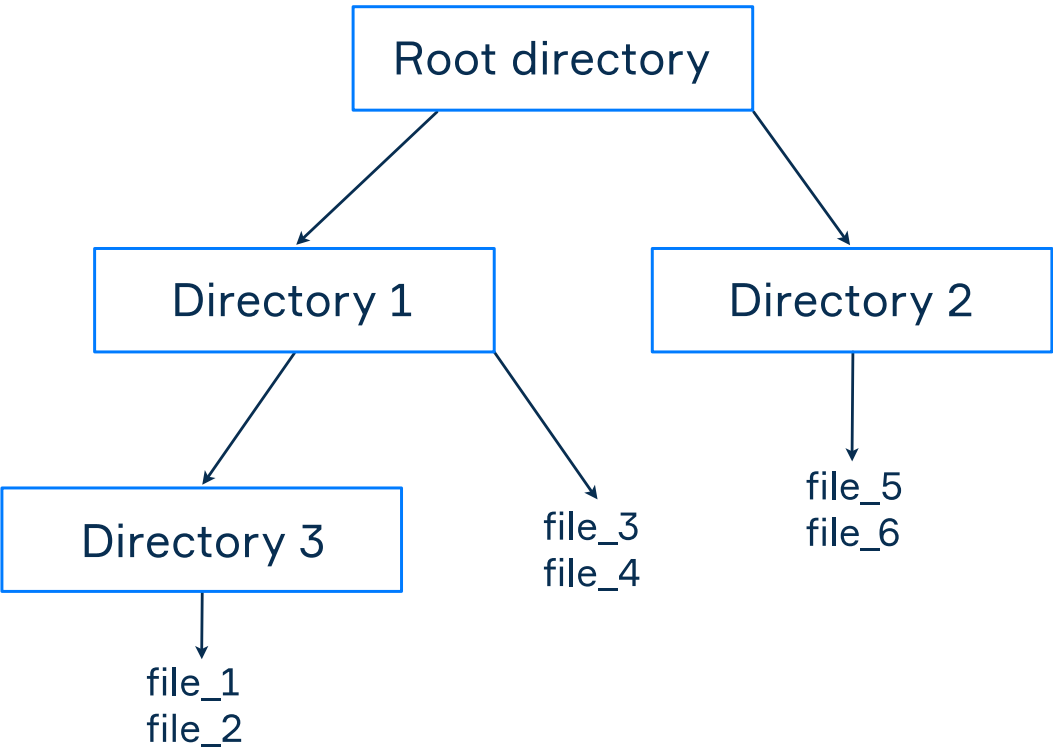
On Windows, root directory usually looks like `C:`, or another drive label, like `D:` . On many UNIXes including MacOs, it looks simply like `/` .

And this is how different data is stored in the computer storage!

🗉 Report a typo

784 users liked this theory. **17** didn't like it. **What about you?**

😍 🙂 😐 🙁 😡

Start practicing

Comments (15)   Hints (0)   Useful links (0)   Show discussion