Technology Module: Tech Skills 201

Extracting Files
Command Line
Virtual Machines

Martin Kong

Expected Outcomes

✓ **Learn how to extract files that are stored in compressed or archive file formats so that they can then be processed in the digital preservation workflow.**

✓ **Learn how to navigate a command line interface and where command line might be used in some of the digital preservation tools.**

✓ **Learn what a virtual machine is and how virtual machines fit into the digital preservation workflow.**

Extracting files from a compressed or storage archive file

What Are Compressed or Storage Archive Files?

On Windows the .zip file format indicates that the file is compressed.

- ➢ It is compressed to save storage space and make transferring files faster.
- ➢ Can also be used on a Mac or Linux.

On Linux the .tar file format indicates that the file is "tape archive" utilized for storing data, and not for compression.

- ➢ They can be compressed after being created, but they become .tgz or .gz.
- ➢ Are just containers for other files, like a folder.
- ➢ Windows cannot handle the extraction of .tar file extension files without the assistance of a third party program such as 7-zip.
- ➢ Can also be used on a Mac.

Compressed or Storage Archive Files…continued

On Mac, the .dmg file format is a Mac OS X Disc Image file, which represents a digital reconstruction of a physical disc.

DMG files are intended for Mac and are often used to store compressed software installers instead of having to use a physical disc.

The Mac operating system is able to handle the decompression and/or extraction of .zip, .tar, and .dmg files natively.

No need for third party software.

Simply double-click on the file to extract it.

Extracting Files…..

….from a compressed archive on Windows (.zip or .tar)

➤ **Windows will natively handle the extraction of zip files.**

➤ **The Windows operating system does not natively handle extracting tar archive files.**

➤ **In order to open a tar file in Windows, a program such as 7-zip _www.7-zip.org_ will have to be installed.** *(screenshots of these steps on the next slide)*

Browse to where the file is located using Windows Explorer.

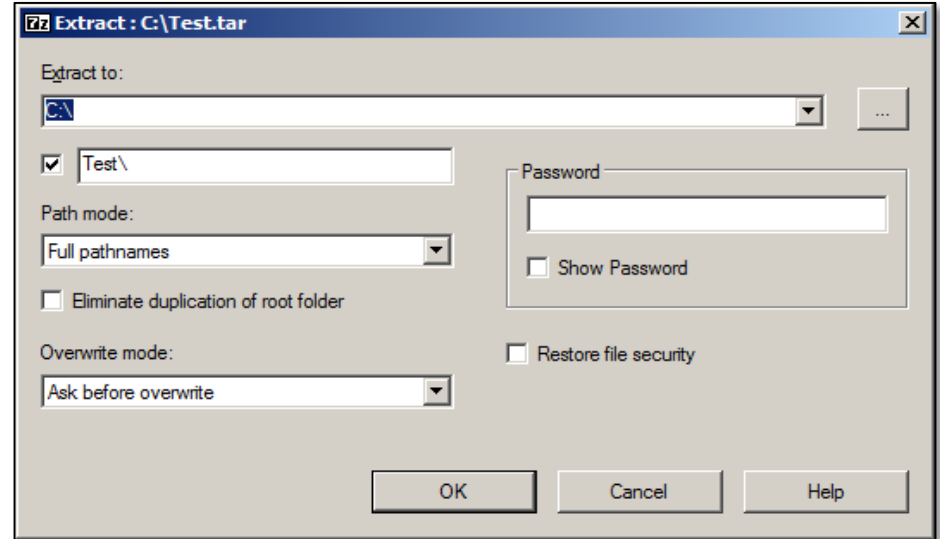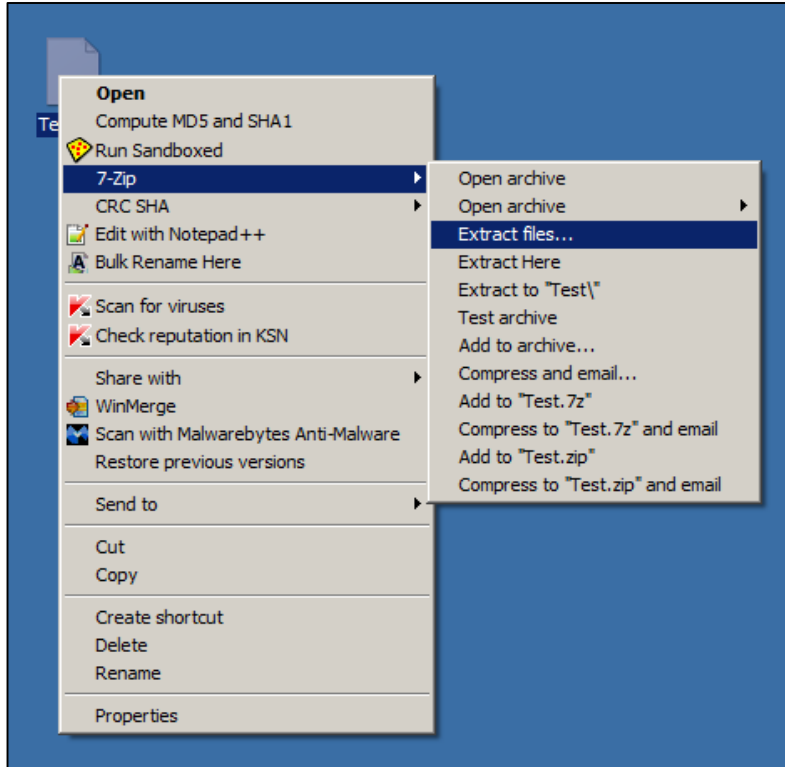Right-click on the tar file and choose the 7-zip submenu and choose "Extract Files…"

This will load the 7-zip program's extraction subprogram window

Choose the directory that you want to extract the files to.

Click OK when you are ready.

The files are now extracted and can be located in the directory you chose.

Extracting Files from a Compressed Archive

Where are compressed or archive files…

….found in digital preservation?

Compressed or archive files can be found in a multitude of places.

The program installers for various digital preservation programs and tools are compressed

➢ They are compressed to save space and reduce time of travel over the Internet.

➢ Must be uncompressed or unpacked before the programs can be installed.

Digital files from donors may also be compressed due to space constraints.

➢ These files will have to be extracted prior to being ingested into the digital preservation workflow.

Navigating a command line interface

What Is The Command Line?

It is different than using a Graphical User Interface (GUI), where you:

Utilize a mouse (or touchscreen) to point and left-or-right click to:

➢ run programs

➢ do work such as web browsing.

The command line is simply another method of interfacing/interacting with a computer.
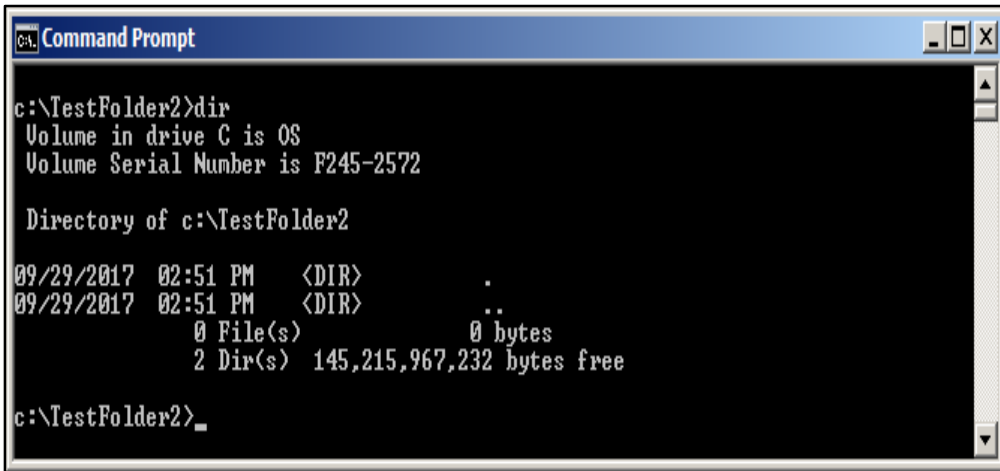
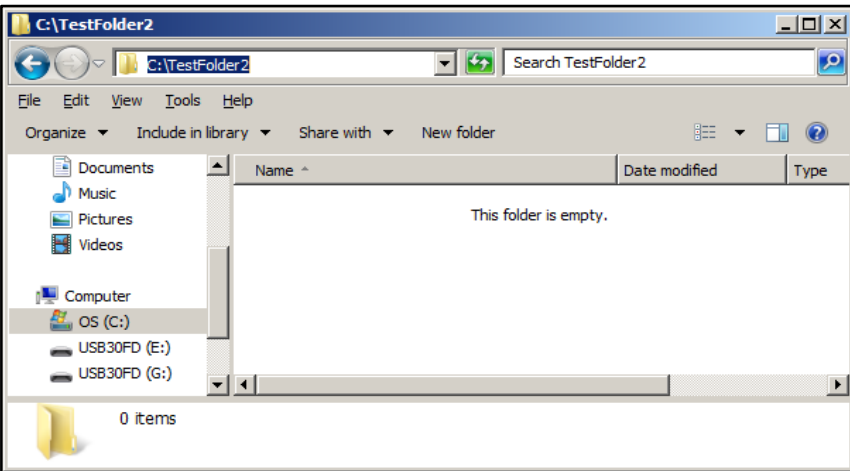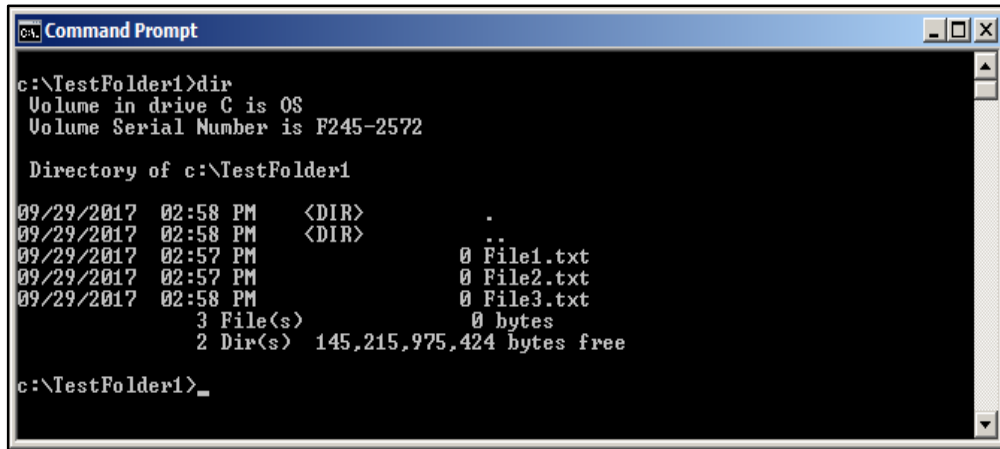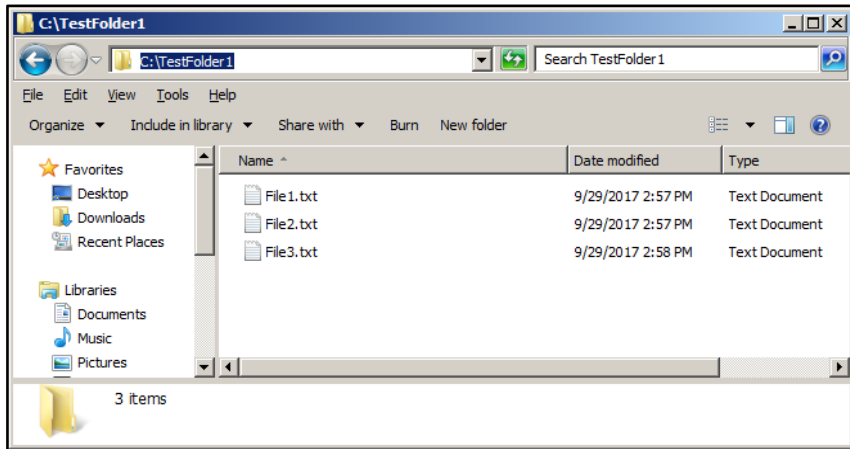**Before the 1980's and 1990's:**

➢ GUI did not yet exist

➢ Command line was the primary method of interacting with computers.
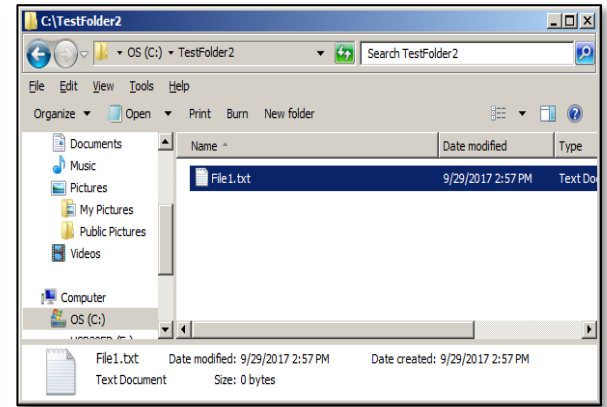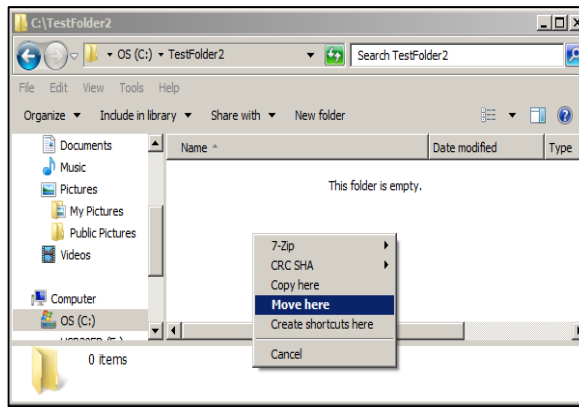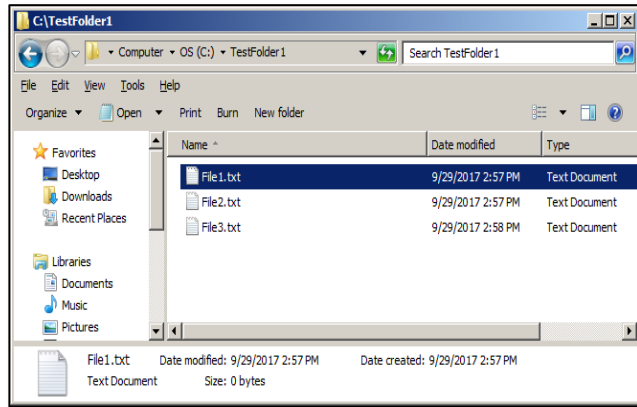
It is textual based

It involves (almost exclusively) the use of the keyboard to

➢ type commands in order to run programs;

➢ copy, move, rename, or delete files;

➢ navigate or browse for files

GUI vs. Command Line

Moving Files in GUI vs. Command Line

How Does The Command Line Work?

Command line is available for use in Mac, Windows, and Unix/Linux computers.
However, we will focus on the command line for Linux computers in the following slides.

Sample prompt after login:

jsmith@computername:~$

The user issues instructions to the system, usually in the form of consecutive lines of textual commands in order to perform an action or actions.

Unlike using a GUI, with a command line, the user is not generally presented with visual commands that can be chosen from.

No right-click menu options to choose commands from.

You are only presented with a blinking cursor

➢ You have to know in advance what it is that you would like to do with a particular file, folder, etc.

➢ For example, in Unix/Linux, if you wanted to rename a file originally called "FILE1.txt" and give it a new name "FILE2.txt", you would type the following into the command line: mv FILE1.txt FILE2.txt

Tips About The Command Line

In Linux, all file or folder/directory names are case sensitive.

/home/jsmith/Docs **is a different directory than** */home/jsmith/docs*

Linux-like emulators are available online

Use to practice your Linux commands as you learn without risk of damage to a real Linux system.

https://autellinux.wordpress.com/2013/09/04/online-simulator-in-linux-practice-linux-commands/

http://www.masswerk.at/jsuix/

This particular emulator seems to work well.

Command Line….Additional Notes

Depending on if you are an administrator or not, you may not have full privileges to browse certain directories in Linux.

➢ You may be able to **navigate** to a specific directory but not **execute** any of the files located in that directory or are not allowed to **make changes** to the files found there.

➢ On the other hand, you could have **full permissions** to make changes to everything in a directory.

➢ In some instances, you don't have permissions to browse above a certain directory other than your home directory i.e. *./home/yourusername*

Where Is The Command Line…..

….in the digital preservation workflow?

Understanding how to use the command line is important for running some of the software tools that are available for use in the digital preservation workflow.

For example, we will be working through an activity that illustrates how to run the File Information Tool Set (FITS)

➢ FITS is a tool that is used to identify, validate and extract technical metadata from digital files.

➢ FITS can handle a multitude of different file types.

A Live Demonstration…..

….of basic LINUX command line commands.

We will be using an online Linux-like emulator for the following demonstration of some basic Linux command line commands

In your packets is a Linux Command Line Cheat Sheet that you can use to obtain additional information or follow along if you would like to do so.

Hands-on Command Line Activity!

Now that we have seen some basic Linux command line commands, we will demonstrate and work together through an activity where you will run the FITS tool.

➢ FITS is used to identify file types for digital files, validates, and extracts technical metadata for digital files.

➢ FITS can be run from the Windows command line

➢ FITS will be run against a very small sample set of different files/file types

➢ We will view the output produced by the program

Virtual Machines
Stress-free Experimentation

Sound Familiar?

➢ "I want to kick the tires on this website software, but there's no demo and I don't have a server to test on."

➢ "Install Linux? UGH."

➢ "I hear this BitCurator thing is all the rage, but it runs on Linux and we only have Windows here."

➢ "I want to learn how to do things on the command line, but I'm scared I'll mess up the server if I do."

TIP: *Never let anyone tell you to type "rm -rf"*

Virtual Machines For The Win!!!!

What's a VM?

➢ **A computer within a computer**

➢ **Can run a completely different operating system**

  For example, the "host" OS can be Windows and the "guest" OS Linux, or vice versa

➢ **Can run server software on a desktop/laptop computer**

➢ **Can create "snapshots" of guest VMs: if something goes wrong, you can go back to a known-good state and try again.**

➢ **Can also destroy a misbehaving guest VM instantly and set up another fast**

  Many popular server OSes have special downloads just for popular VM software; you don't even have to do most of the installation work!

VM Software: Hypervisor

Popular: VirtualBox and VMWare

➢ VirtualBox is open source. VMWare has a free version called VMWare Player that's usually good enough.

Both are fine. It's personal preference.

Cloud VM services available inexpensively (not free!)

➢ DigitalOcean.com is probably the easiest to use.

➢ Handy for places with strict rules about installing software on staff computers, or very low-powered computers

Installing a Guest OS

If you're lucky, there will be pre-installed VMs for your hypervisor of choice.

If you're not as lucky, there will be an "ISO" (CD image) to install from.

➢ This is a little tricky, but once you get the hang of it you'll be fine. VirtualBox tends to be a bit easier than VMWare.

➢ BitCurator comes as an ISO.

As soon as you have a working install, shut it down and take a snapshot!

➢ That way you can always return to it if something goes wrong.

Virtual Machines in Digital Preservation

Virtual Machines can be useful in Digital Preservation

If you don't have access to another physical computer and you want to run or test digital preservation software that runs in another OS type such as Linux.

- ➢ BitCurator
- ➢ Archivematica

You don't want to install or are restricted from installing software directly on your actual computer.

- ➢ IT approval or assistance may still be needed to install the virtual machine software or hypervisor on your local machine.
- ➢ Once VM software is installed, then you can run as many virtual machines as needed.

Please Standby….

At the end of this session, if time allows, a demonstration will be given of a virtual machine running the Linux operating system.

*Linux will be running **INSIDE** of the Virtual Machine.*

Technology Module: Tech Skills 201

Extracting Files
Command Line
Virtual Machines

QUESTIONS?