

Sesión 1: ¿Con qué trabajamos?

Carlos Ugarte

28.08.2025

1 Introducción

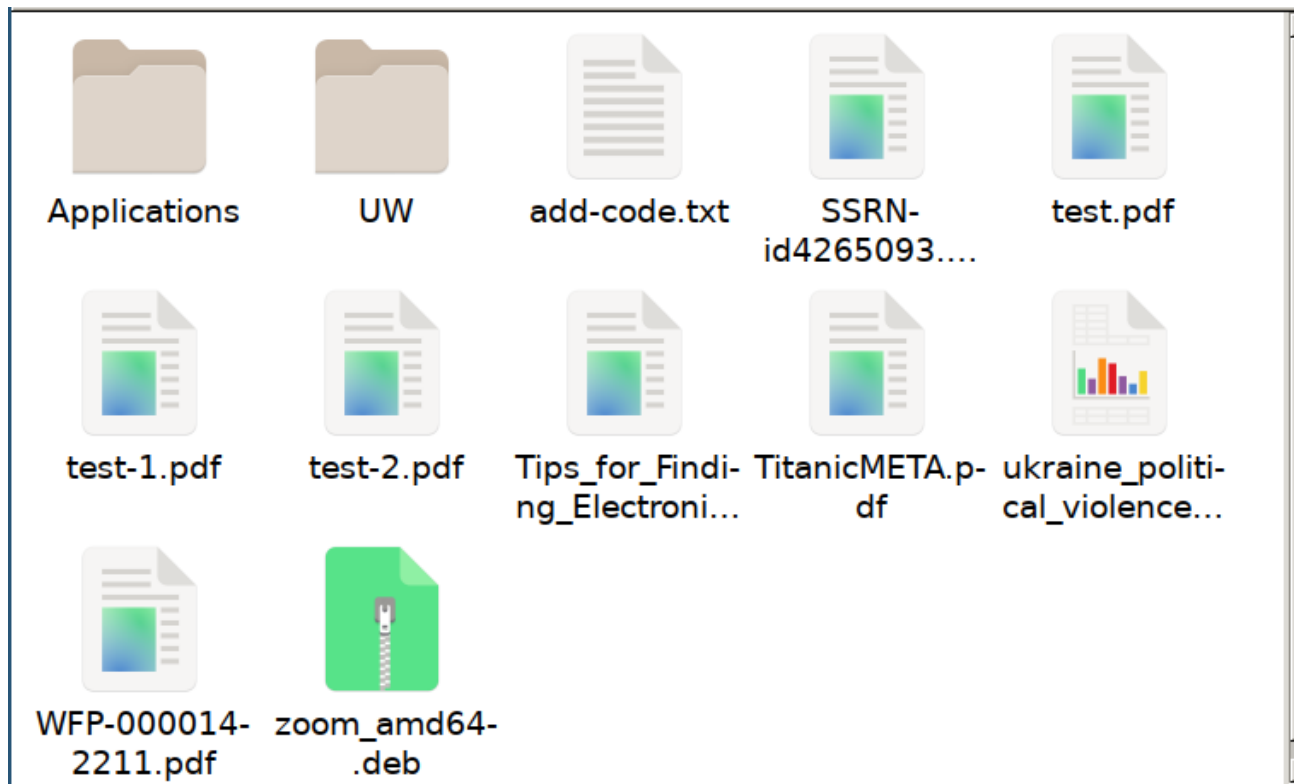
La lingüística computacional requiere trabajar con una gran cantidad de archivos: scripts con código, documentos, tareas, archivos de datos, etc. Por lo tanto, es necesario saber cómo gestionarlos. En esta sesión, hablaremos sobre cómo organizar nuestros archivos y cómo concebir esta organización como un árbol de archivos. Aunque no sea común pensar en los archivos de una computadora como un árbol, así es como los conciben las computadoras y los lenguajes de programación. Tener una visión general de la distribución del sistema de archivos hará más fácil decirle a nuestro código cómo acceder a ellos.

2 ¿Cómo encontrar tus archivos?

Para poder cargar nuestros archivos, discutiremos dos conceptos importantes: árbol del sistema de archivos (file system tree) y directorio de trabajo (working directory).

Las computadoras almacenan sus archivos y carpetas como un árbol. En esta analogía, las carpetas serían como ramas que pueden ramificarse aún más (contener otras carpetas), y los archivos como hojas, los extremos de las ramas.

Veamos un ejemplo:





amath352



cse142



info201



exercises



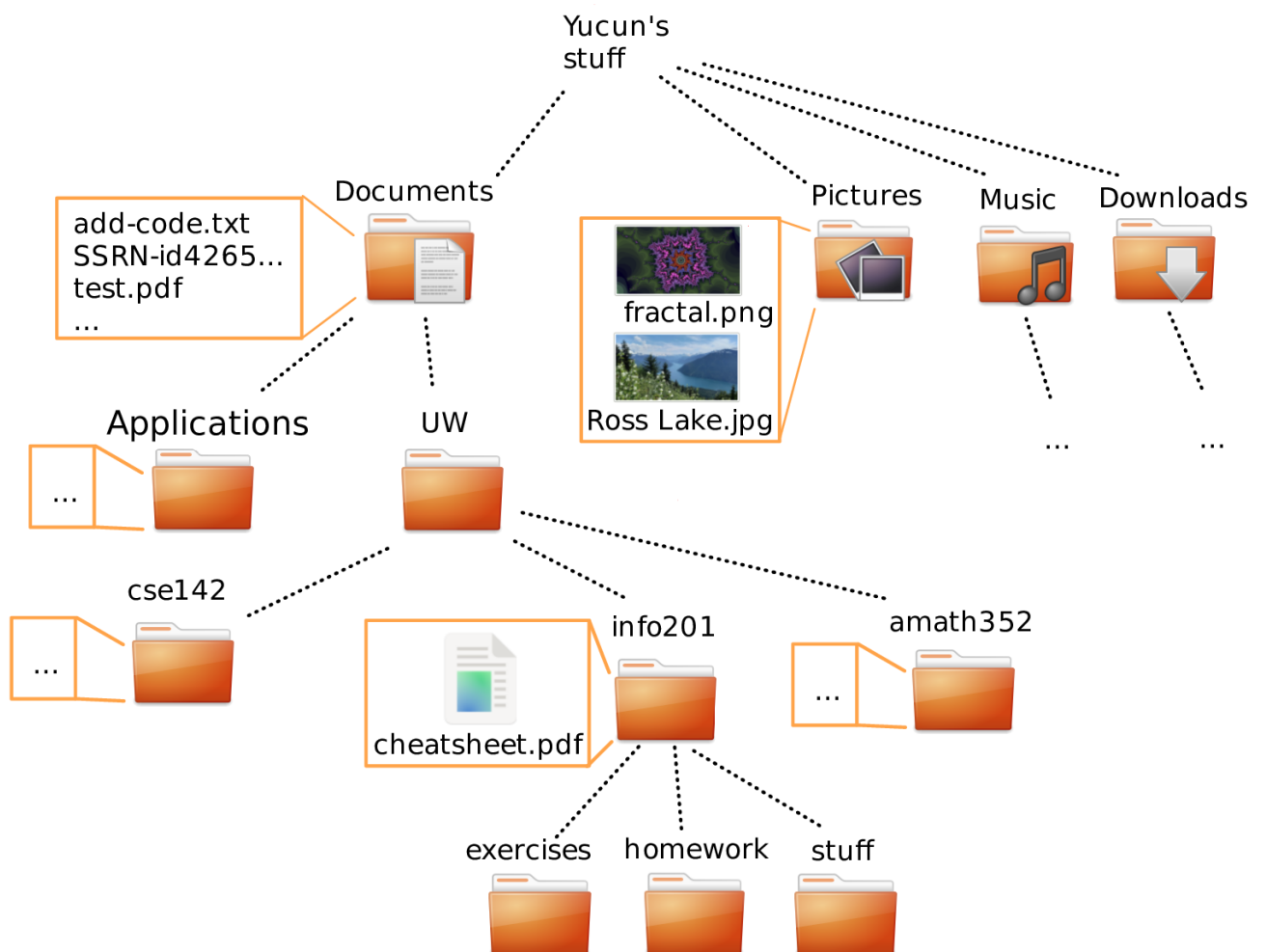
homework



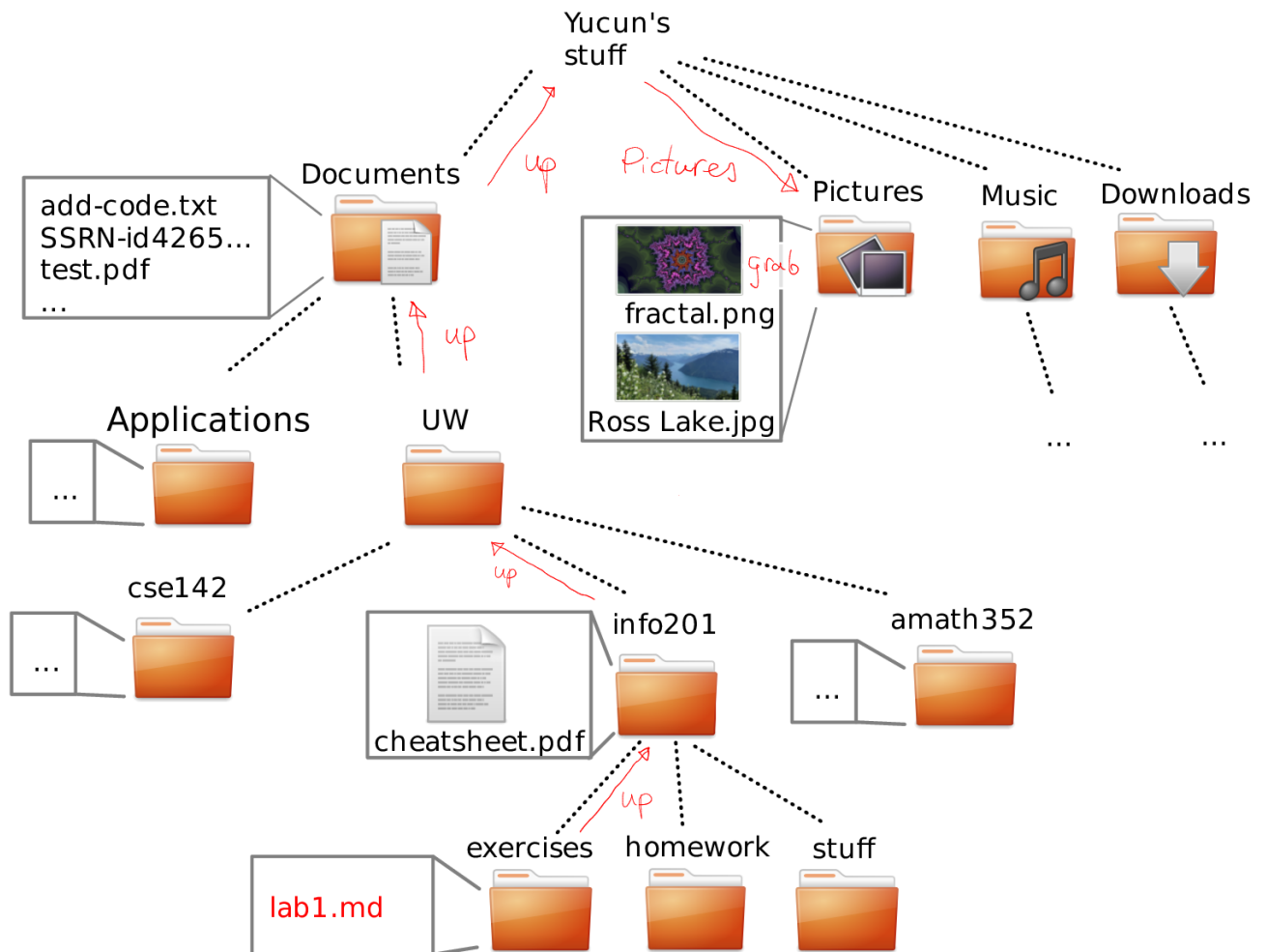
stuff



cheatsheet.-
pdf



Esto es lo que llamamos el árbol de sistema de archivos (file system tree). Los programas "piensan" que están "situados" en un lugar en el file system tree en TODO momento. El lugar en el cual están situados es lo que llamamos directorio de trabajo (working directory). Este es un concepto importante porque siempre que un programa intente cargar archivos, empezará a hacerlo desde donde esté "parado".



Esto involucra una serie de pasos:

1. arriba (adentro de INFO201)
2. arriba (adentro de UW)
3. arriba (adentro de DOCUMENTS)
4. arriba (adentro de YUCUN'S STUFF)
5. adentro de PICTURES
6. agarrar FRACTAL.PNG de ahí

Al ir arriba, no se necesita especificar el nombre de la carpeta ya que solo hay una manera de ir hacia arriba. En cambio, al ir abajo, es necesario especificar el lugar al que uno quiere bajar. Por eso el paso 5 incorpora el nombre de la carpeta a la que nos dirigimos.

Esta serie de pasos se parece a las indicaciones de navegación en aplicaciones de navegación, como google maps. En la computadora, a esta serie de pasos se le conoce como la ruta relativa (relative path). Este explica cómo llegar a algún lugar desde la posición actual (working directory), en relación con la posición actual.

Aunque intuitivo para nosotros, la computadora necesita una manera más compacta de recibir las instrucciones. Los cambios serían:

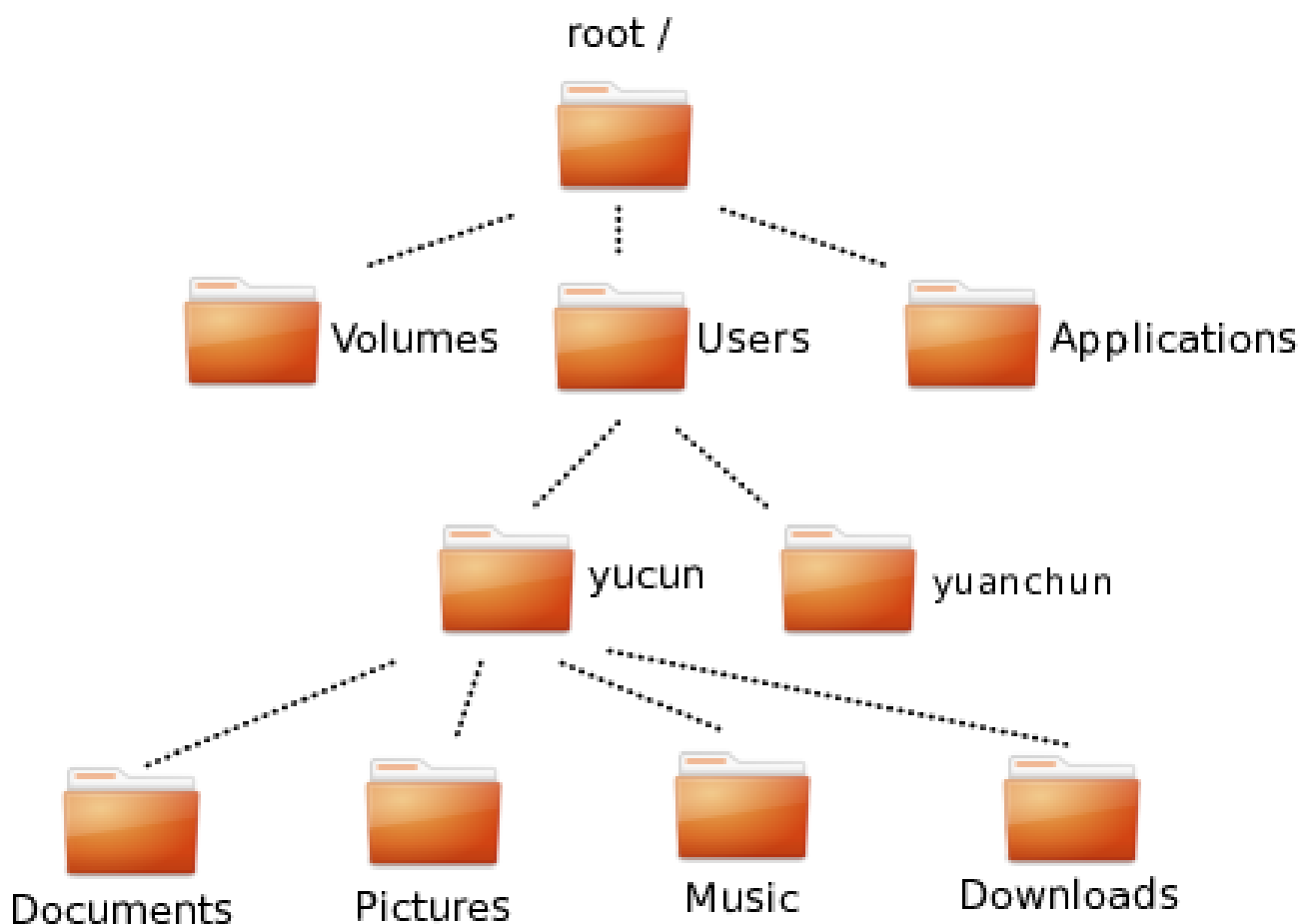
- remplazar “arriba” con dos puntos “..”
- agregar un slash “/” entre las instrucciones en lugar de crear una nueva
- llamar al archivo simplemente por su nombre “fractal.png”

Por lo que, para la computadora, sería mejor ver un relative path como el siguiente:

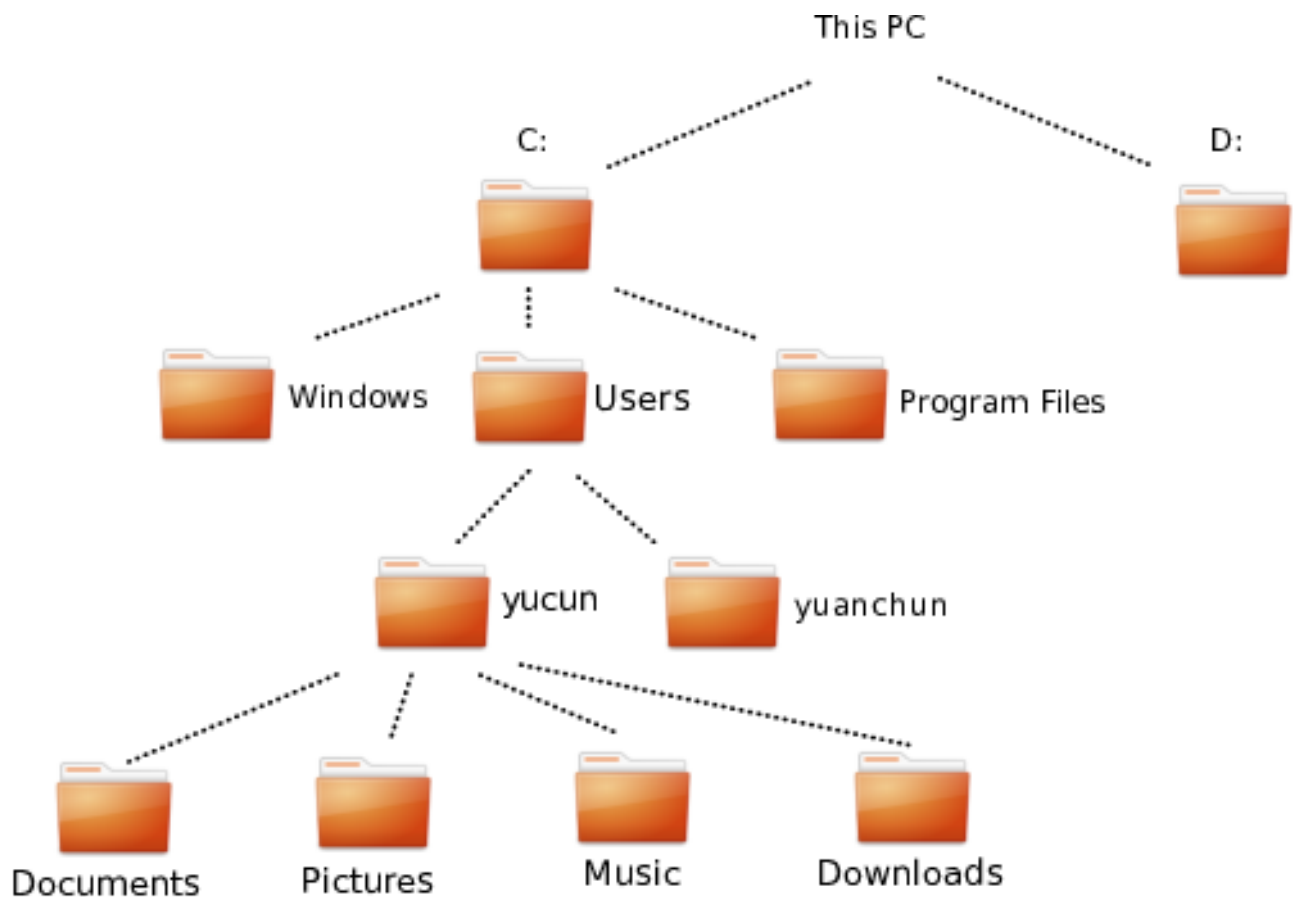
```
../../../../Picture/fractal.png
```

Pero hay mucho más. La imagen anterior solo muestra una rama de todo el árbol. Tu computadora necesita muchos de archivos, data, apps para correr. Asimismo, las computadoras pueden tener más de un usuario (no solo Yucun, en nuestro ejemplo).

Entonces, todo el file system tree lucería algo así, si es que uno tiene Mac:



O de la siguiente manera, si es que uno tiene PC:



El folder de Yucun (que recibe el nombre de tu usuario en tu computadora) comprende todos tus archivos. Este es llamado el HOME FOLDER o HOME DIRECTORY. Entonces si queremos hacer la misma lista de direcciones, pero de forma compacta, quedaría de la siguiente forma.

Para Mac:

```
/Users/yucun/Documents/info201/cheatsheet.pdf
```

Para PC:

```
C:/Users/yucun/Documents/UW/info201/cheatsheet.pdf
```

Estas reglas de navegación son llamadas la ruta absoluta (absolute path). La ventaja de la ruta absoluta es que siempre dice dónde está el archivo ubicado, sin importar el directorio de trabajo actual. Sin embargo, si uno mueve el archivo de lugar, o le cambia de nombre, entonces la ruta absoluta ya no es válida.

Entonces, existen dos maneras de navegar por el sistema de archivos: ruta relativa y ruta absoluta. La ruta relativa comienza con el directorio de trabajo actual, y el primer paso suele ser ascender en el árbol, antes de descender en otras carpetas. La ruta absoluta comienza con la carpeta raíz, la carpeta superior, y no es necesario ascender de nuevo después de descender en un directorio.

Las rutas absolutas y relativas son similares, ya que ambas permiten acceder a archivos y directorios individuales en el equipo. La única diferencia es que la ruta relativa comienza desde el directorio de trabajo actual, mientras que la ruta absoluta comienza desde la carpeta raíz.

3 ¿Cómo organizar tu computadora?

La organización de tus archivos depende únicamente de ti. Lo importante al respecto es que uno organice sus archivos conscientemente. Este taller requiere que crees folders, cargues y guardes archivos. Muchas veces pasa que no sabemos dónde terminan nuestros archivos resultantes al trabajar con métodos computacionales.

Mi recomendación:

Tener todas tus carpetas de proyectos dentro de un directorio llamado: PROYECTOS en la carpeta de DOCUMENTOS en tu computadora. A lo largo del taller, veremos que esto es beneficioso porque en la misma carpeta DOCUMENTOS_PROYECTOS trabajaremos otros conceptos tales como los FRESH VIRTUAL ENVIRONMENTS y las carpetas de repositorios. Eso lo haremos ahora, discutiendo sobre la noción de command line.

4 Command line

La línea de comandos (command line) es una interfaz para una computadora: una forma de comunicarse con la máquina. Pero a diferencia de las interfaces gráficas comunes que usan ventanas, iconos, menús y punteros, la línea de comandos se basa en texto: se escriben comandos en lugar de hacer clic en iconos. La línea de comandos permite hacer todo lo que normalmente harías con un clic del ratón, pero escribiendo de forma similar a la programación.

Para acceder a la terminal desde Windows, presiona la tecla de Windows + R, escribe cmd y presiona Enter para abrir el Símbolo del sistema.

Ahora una lista de comandos para el manejo de archivos en tu computadora:

Acción	Sintaxis
Cambiar de directorio	cd PATH
Crear nuevo directorio	mkdir DIRECTORIO_NUEVO
Lista de contenidos en un directorio	ls
Copiar un archivo	cp ARCHIVO_VIEJO ARCHIVO_NUEVO
Mover un archivo	mv ARCHIVO_VIEJO ARCHIVO_NUEVO
Eliminar un archivo	rm NOMBRE_ARCHIVO
Crear un archivo nuevo	echo. > NOMBRE_ARCHIVO
Abrir un archivo	start NOMBRE_ARCHIVO

Entonces, ahora crearemos con esto comandos un directorio nuevo en nuestras computadoras que tenga el siguiente absolute path:

Para Mac:

```
/Users/usuario/Documentos/Proyectos/taller_pucp25
```

Para PC:

```
C:/Users/usuario/Documentos/Proyectos/taller_pucp25
```

En la command line se especifica el último directorio en el que uno se encuentra (working directory). Una manera sencilla de saber con más precisión toda la ruta del working directory es con el comando PWD.

Una vez creado el directorio, clonaremos el repositorio del taller que se encuentra en Github a nuestras computadoras. Todo el material del curso lo encontraremos ahí.

Para aquello, seguimos los siguientes pasos:

1. Instalar Python
<https://www.python.org/downloads/>
2. Instalar git
<https://git-scm.com/downloads/>

Luego, nos “situamos” en la carpeta TALLER_PUCP25. Ejecutamos el siguiente comando para clonar el repositorio:

```
git clone https://github.com/MuffinLinwist/taller_pucp25.git
```

5 Actividad para la casa

Actividad correspondiente a la primera sesión: Comienza tu proyecto

1. Crea un repositorio desde la terminal en tu computadora para tu nuevo proyecto en la siguiente ruta: C:/Usuarios/x/Documentos/Proyectos/mi_proyecto.
2. Dentro de ese repositorio, crea un documento README.md que cuente tus intereses o dudas sobre la lingüística computacional.
3. Lee un poco acerca de Github y anota preguntas en el README.md.