GenEpi-BioTrain – Virtual Training 9 & 10

Unix for beginners: Combining Files

August 21st 2024

Additional Information

- Use the **Handout**! Everything we are presenting is already in the Handout.

- If you are a mac user and had trouble using `wget` and `readlink` yesterday, read the **macOS addendum** on EVA. Here you can find installation guides and alternative commands.

Learning goals

Today, we will learn how to:

- Combine files
- Writing and appending to files
- Searching within files and folders
- Counting words/lines
- Redirecting output from one program to another using |
- Compressing and decompressing files
- Creating empty files

Objectives

Specific objectives of this session:

- Redirect & Append Output to File
- Combining Files

Objectives

Specific objectives of this session:

- **Redirect & Append Output to File**
- Combining Files

Redirecting `command > file`

Direct the output of a command to a file

```
ls > contentlist.txt
```

```
/
├── home
│   └── alice
│       ├── Documents
│       │   ├── resume.docx
│       │   ├── project_notes.txt
│       │   └── work ←
│       │       ├── report1.pdf
│       │       ├── presentation.pptx
│       │       ├── timesheet.xlsx
│       │       └── private
│       ├── Downloads
│       └── ...
├── etc
└── ...
```

contentlist.txt
report1.pdf
presentation.pptx
timesheet.xlsx
private

Appending `command >> file`

Append the output of a command to a file

```
ls >> contentlist.txt
```

```
/
├── home
│   └── alice
│       ├── Documents ←
│       │   ├── resume.docx
│       │   ├── project_notes.txt
│       │   └── work
│       │       ├── report1.pdf
│       │       ├── presentation.pptx
│       │       ├── timesheet.xlsx
│       │       └── private
│       ├── Downloads
│       └── ...
├── etc
└── ...
```

→

```
contentlist.txt
---------------------
report1.pdf

presentation.pptx

timesheet.xlsx

private

resume.docx

project_notes.txt
```

Objectives

Specific objectives of this session:

- Redirect & Append Output to File
- **Combining Files**

Combining Files

`cat file1 file2 > file3`

concatenate multiple files into one

```
cat contentlist.txt movielist.txt > merged.txt
```

contentlist.txt
report1.pdf
presentation.pptx
timesheet.xlsx
private
resume.docx
project_notes.txt

movielist.txt
The Godfather (1972)
Inception (2010)
The Shawshank Redemption (1994)

merged.txt
report1.pdf
presentation.pptx
timesheet.xlsx
private
resume.docx
project_notes.txt
The Godfather (1972)
Inception (2010)
The Shawshank Redemption (1994)

Exploration

Exercise 1

Checkpoint: 3

Have fun! ☺

EUROPEAN CENTRE FOR DISEASE PREVENTION AND CONTROL

GenEpi-BioTrain – Virtual Training 9 & 10

Unix for beginners: **Searching within Files**

August 21st 2024

Objectives

Specific objectives of this session:

- grep: Searching within Files
- Introduction to FASTA Files

Objectives

Specific objectives of this session:

- **grep: Searching within Files**
- Introduction to FASTA Files

Searching

`grep [options] "pattern" file`

Searching text or files for lines that match a specified `pattern`

```
grep "The" merged.txt
```

merged.txt

```
report1.pdf
presentation.pptx
timesheet.xlsx
private
resume.docx
project_notes.txt
The Godfather (1972)
Inception (2010)
The Shawshank Redemption (1994)
```

→

```
The Godfather (1972)
The Shawshank Redemption (1994)
```

Searching

`grep [options] "pattern" file`

Searching text or files for lines that match a specified `pattern`

```
grep "the" merged.txt
```

merged.txt
report1.pdf
presentation.pptx
timesheet.xlsx
private
resume.docx
project_notes.txt
The Godfather (1972)
Inception (2010)
The Shawshank Redemption (1994)

→

Searching

`grep [options] "pattern" file`

Searching text or files for lines that match a specified `pattern`

```
grep -i "the" merged.txt
```

```
merged.txt
report1.pdf
presentation.pptx
timesheet.xlsx
private
resume.docx
project_notes.txt
The Godfather (1972)
Inception (2010)
The Shawshank Redemption (1994)
```

→

```
The Godfather (1972)
The Shawshank Redemption (1994)
```

Searching

`grep [options] "pattern" file`

Searching text or files for lines that match a specified pattern

```
grep -in "the" merged.txt
```

merged.txt
report1.pdf
presentation.pptx
timesheet.xlsx
private
resume.docx
project_notes.txt
The Godfather (1972)
Inception (2010)
The Shawshank Redemption (1994)

→

```
7:The Godfather (1972)
9:The Shawshank Redemption (1994)
```

Searching

`grep [options] "pattern" file`

Searching text or files for lines that match a specified `pattern`

```
grep -io "the" merged.txt
```

merged.txt

```
report1.pdf
presentation.pptx
timesheet.xlsx
private
resume.docx
project_notes.txt
The Godfather (1972)
Inception (2010)
The Shawshank Redemption (1994)
```

→

```
The
The
```

Objectives

Specific objectives of this session:

- grep: Searching within Files
- **Introducing FASTA Files**

FASTA File

A FASTA file is a text format for storing biological sequences like DNA, RNA, or proteins.

example.fasta

```
>sequence_name1 [description]
ATGCGTACGTAGCTAGCCGTAGCTAGCTAGTACG
>sequence_name2 [description]
TTGCACATAAATGTTGTCTATCTCTAATGT
```

Header: Sequence name and optional description. Always starts with ">"

Sequence: Nucleotide or protein sequence

Exploration

Exercise 2

Checkpoint: 4

Have fun! ☺

GenEpi-BioTrain – Virtual Training 9 & 10

Unix for beginners: Redirecting output

August 21st 2024

Objectives

Specific objectives of this session:

- Counting
- Redirecting Output to Command

Objectives

Specific objectives of this session:

- **Counting**
- Redirecting Output to Command

Word count

```
wc [options] [file]
```

Counts the number of lines, words, and characters in a file or input

```
wc contentlist.txt
```

contentlist.txt

```
report1.pdf
presentation.pptx
timesheet.xlsx
private
resume.docx
project_notes.txt
The Godfather (1972)
Inception (2010)
The Shawshank Redemption (1994)
```

→

```
9      16      157          contentlist.txt
```

Word count

```
wc [options] [file]
```

Counts the number of lines, words, and characters in a file or input

```
wc contentlist.txt
```

contentlist.txt

```
report1.pdf
presentation.pptx
timesheet.xlsx
private
resume.docx
project_notes.txt
The Godfather (1972)
Inception (2010)
The Shawshank Redemption (1994)
```

```
Lines  Words  Characters  File name
9      16     157         contentlist.txt
```

Word count

`wc [options] [file]`

Counts the number of lines, words, and characters in a file or input

```
wc -l contentlist.txt
```

```
contentlist.txt
report1.pdf
presentation.pptx
timesheet.xlsx
private
resume.docx
project_notes.txt
The Godfather (1972)
Inception (2010)
The Shawshank Redemption (1994)
```

→

```
Lines  File name
9      contentlist.txt
```

Word count

```
wc [options] [file]
```

Counts the number of lines, words, and characters in a file or input

```
wc -c contentlist.txt
```

```
contentlist.txt
report1.pdf
presentation.pptx
timesheet.xlsx
private
resume.docx
project_notes.txt
The Godfather (1972)
Inception (2010)
The Shawshank Redemption (1994)
```

→

```
Characters  File name
157         contentlist.txt
```

Objectives

Specific objectives of this session:

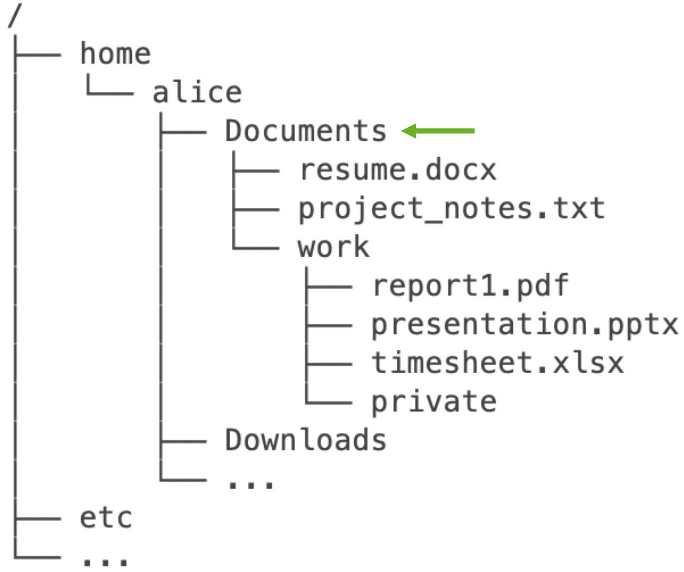
- Counting
- **Redirecting Output to Command**

Pipe

```
command | command
```

Takes the output of one command and use it as the input for another command

```
ls -l | grep "Apr"
```



```
/
├── home
│   └── alice
│       ├── Documents
│       │   ├── resume.docx
│       │   ├── project_notes.txt
│       │   └── work
│       │       ├── report1.pdf
│       │       ├── presentation.pptx
│       │       ├── timesheet.xlsx
│       │       └── private
│       ├── Downloads
│       └── ...
├── etc
└── ...
```

Pipe

`command | command`

Takes the output of one command and use it as the input for another command

```
ls -l | grep "Apr"
```

```
/
├── home
│   └── alice
│       ├── Documents ←
│       │   ├── resume.docx
│       │   ├── project_notes.txt
│       │   └── work
│       │       ├── report1.pdf
│       │       ├── presentation.pptx
│       │       ├── timesheet.xlsx
│       │       └── private
│       ├── Downloads
│       └── ...
├── etc
└── ...
```

→

```
-rw-r--r--  1  alice  53579   May 13  15:22  resume.docx
-rw-r--r--  1  alice  157017  Apr 6   07:27  project_notes.txt
drwxr-xr-x  1  alice  736     Apr 23  09:02  work
```

Pipe

```
command | command
```

Takes the output of one command and use it as the input for another command

```
ls -l | grep "Apr"
```

```
/
├── home
│   └── alice
│       ├── Documents
│       │   ├── resume.docx
│       │   ├── project_notes.txt
│       │   └── work
│       │       ├── report1.pdf
│       │       ├── presentation.pptx
│       │       ├── timesheet.xlsx
│       │       └── private
│       ├── Downloads
│       └── ...
├── etc
└── ...
```

```
-rw-r--r--   1  alice   53579    May 13   15:22   resume.docx
-rw-r--r--   1  alice   157017   Apr 6    07:27   project_notes.txt
drwxr-xr-x   1  alice   736      Apr 23   09:02   work
```

```
-rw-r--r--   1  alice   157017   Apr 6    07:27   project_notes.txt
drwxr-xr-x   1  alice   736      Apr 23   09:02   work
```

Exploration

Exercise 3

Checkpoint: 4

Have fun! ☺

GenEpi-BioTrain – Virtual Training 9 & 10

Unix for beginners: Putting it all together

August 21st 2024

Objectives

Specific objectives of this session:

- Compressing and Decompressing Files

Compressing

```
tar [options] file target
```

Compress directory to zip-file

```
tar -czvf work.tar.gz ~/Documents/work/
```

Option	Description
`-c`	**C**reate a new archive
`-x`	E**x**tract files from the archive
`-z`	Compress the archive using g**z**ip
`-v`	**V**erbosely list the files being processed
`-f`	Specify the name of the archive **f**ile
`-h`	Follow symbolic links and include the actual files they point to.

Decompressing `tar [options] file`

Decompress `zip-file`

```
tar -xzvf work.tar.gz
```

Option	**Description**
`-c`	**C**reate a new archive
`-x`	E**x**tract files from the archive
`-z`	Compress the archive using g**z**ip
`-v`	**V**erbosely list the files being processed
`-f`	Specify the name of the archive **f**ile
`-h`	Follow symbolic links and include the actual files they point to.

Exploration

Exercise 4

Checkpoint: 4

Have fun! ☺

GenEpi-BioTrain – Virtual Training 9 & 10

Unix for beginners: Wrapping up

August 21st 2024

Objectives

Specific objectives of this session:

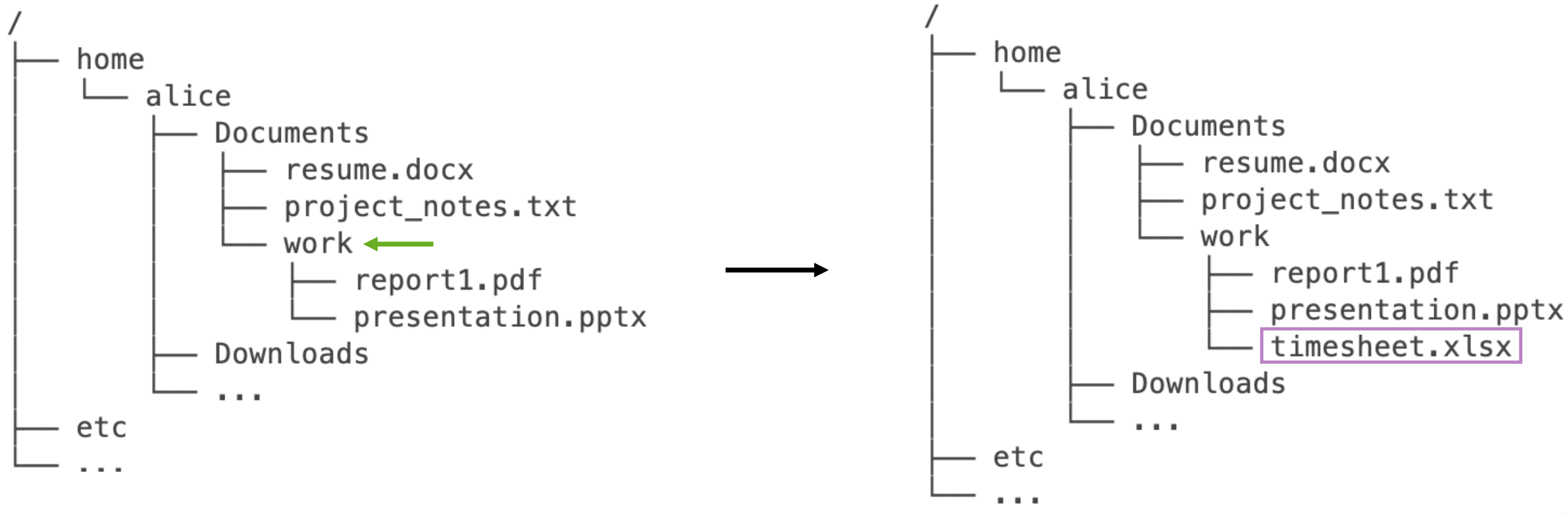
- Creating files ☺ easy – you are experts now!

Create File

`touch [options] file`

Create new empty file

`touch timesheet.xlsx`



```
/
├── home
│   └── alice
│       ├── Documents
│       │   ├── resume.docx
│       │   ├── project_notes.txt
│       │   └── work
│       │       ├── report1.pdf
│       │       └── presentation.pptx
│       ├── Downloads
│       └── ...
├── etc
└── ...
```

```
/
├── home
│   └── alice
│       ├── Documents
│       │   ├── resume.docx
│       │   ├── project_notes.txt
│       │   └── work
│       │       ├── report1.pdf
│       │       ├── presentation.pptx
│       │       └── timesheet.xlsx
│       ├── Downloads
│       └── ...
├── etc
└── ...
```

Exploration

Exercise 5

Checkpoint: 5

Have fun! ☺

We will meet up at 13:55

Rounding up – Day 2

Today, we learned how to:

- Combining files
- Writing and appending to files
- Searching within files and folders
- Counting words/lines
- Redirecting output from one program to another using |
- Compressing and decompressing files
- Creating empty files

Thanks for coming

Great job! ☺

Tomorrow we will try out the Conda ecosystem and how to mange different environments