Introduction to UNIX

Kamil S Jaron, Marc Robinson-Rechavi

22.9. 2016

Why UNIX?

Sequencing reads of 15 genomes

$\approx$ 1GB of plain text data / genome

$\approx$ 640938 of pages

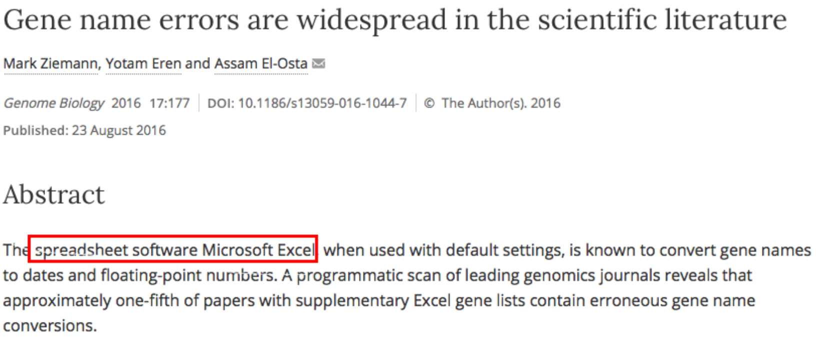
How to check a file??

Notepad?

not well suited for big files

Office?

Use at least read-only mode



Gene name errors are widespread in the scientific literature

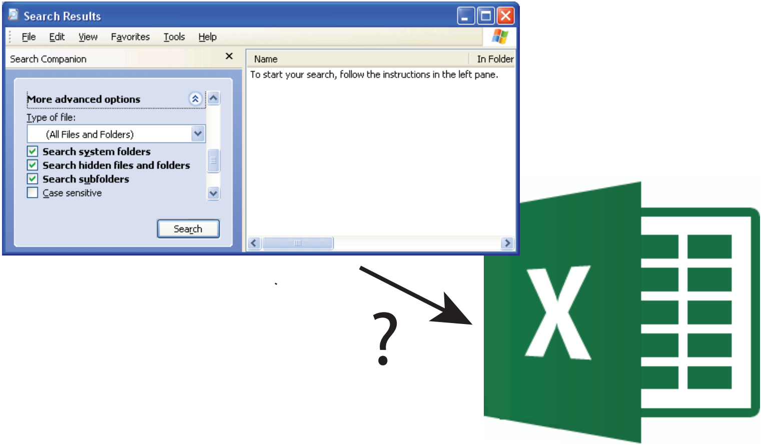
Mark Ziemann, Yotam Eren and Assam El-Osta

Genome Biology 2016 17:177 | DOI: 10.1186/s13059-016-1044-7 | © The Author(s). 2016

Published: 23 August 2016

Abstract

The spreadsheet software Microsoft Excel when used with default settings, is known to convert gene names to dates and floating-point numbers. A programmatic scan of leading genomics journals reveals that approximately one-fifth of papers with supplementary Excel gene lists contain erroneous gene name conversions.



More advanced options

Type of file: (All Files and Folders)

☑ Search system folders

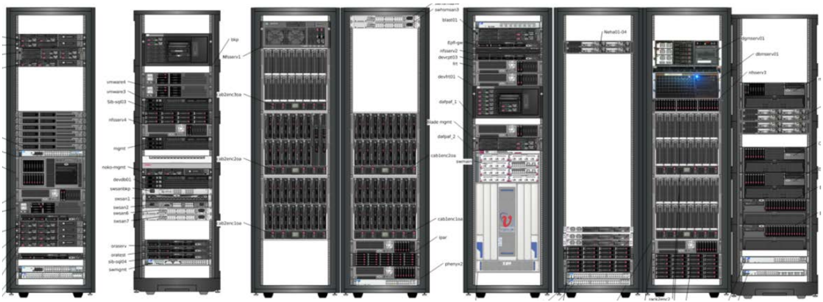
☑ Search hidden files and folders

☑ Search subfolders

☐ Case sensitive

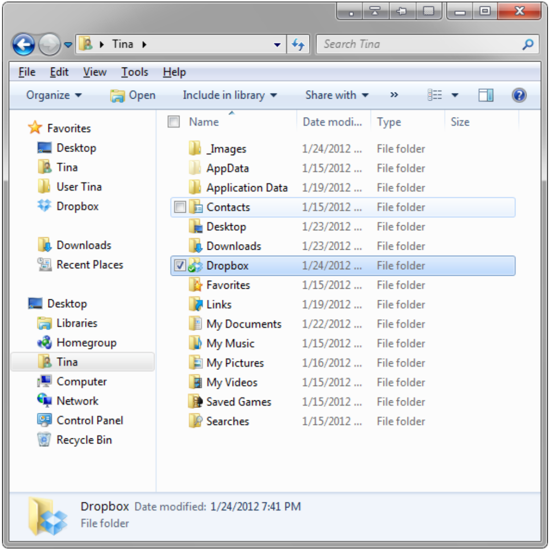
To start your search, follow the instructions in the left pane.

Cluster computing



How can UNIX help us?

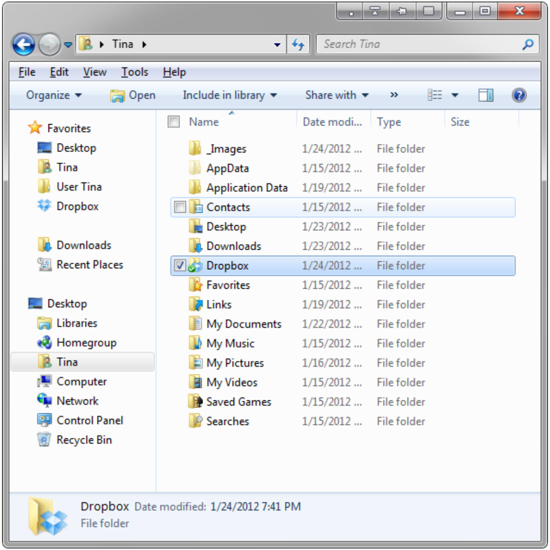
Command line ≈ explorer + toolbox of commands



	Name	Date modi...	Type	Size
	_Images	1/24/2012 ...	File folder	
	AppData	1/15/2012 ...	File folder	
	Application Data	1/19/2012 ...	File folder	
□	Contacts	1/15/2012 ...	File folder	
	Desktop	1/23/2012 ...	File folder	
	Downloads	1/23/2012 ...	File folder	
☑	Dropbox	1/24/2012 ...	File folder	
	Favorites	1/15/2012 ...	File folder	
	Links	1/19/2012 ...	File folder	
	My Documents	1/22/2012 ...	File folder	
	My Music	1/15/2012 ...	File folder	
	My Pictures	1/16/2012 ...	File folder	
	My Videos	1/15/2012 ...	File folder	
	Saved Games	1/15/2012 ...	File folder	
	Searches	1/15/2012 ...	File folder	

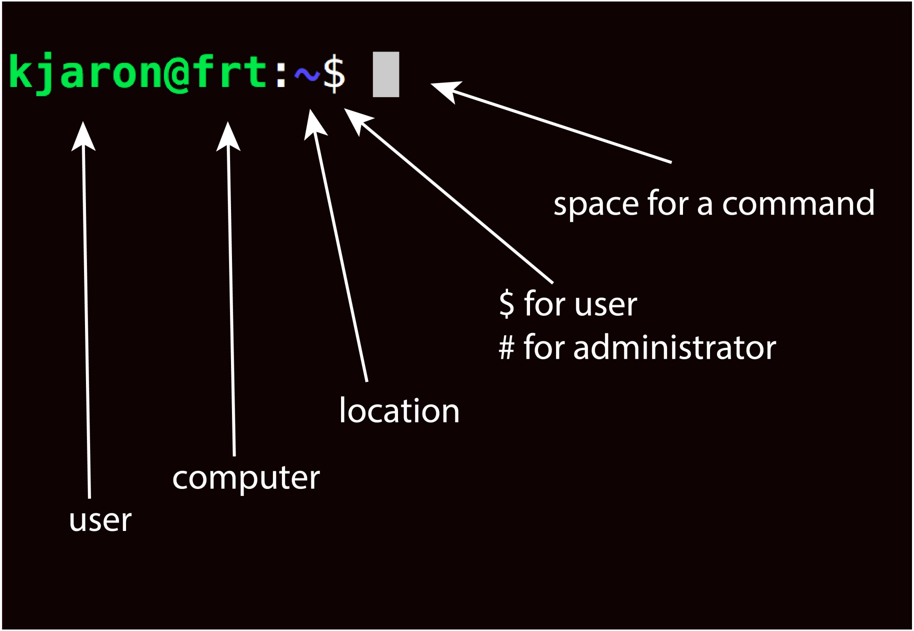
Dropbox Date modified: 1/24/2012 7:41 PM
File folder

bash ≈ explorer



□ Name	Date modi...	Type	Size
_Images	1/24/2012 ...	File folder	
AppData	1/15/2012 ...	File folder	
Application Data	1/19/2012 ...	File folder	
□ Contacts	1/15/2012 ...	File folder	
Desktop	1/23/2012 ...	File folder	
Downloads	1/23/2012 ...	File folder	
☑ Dropbox	1/24/2012 ...	File folder	
Favorites	1/15/2012 ...	File folder	
Links	1/19/2012 ...	File folder	
My Documents	1/22/2012 ...	File folder	
My Music	1/15/2012 ...	File folder	
My Pictures	1/16/2012 ...	File folder	
My Videos	1/15/2012 ...	File folder	
Saved Games	1/15/2012 ...	File folder	
Searches	1/15/2012 ...	File folder	

Dropbox Date modified: 1/24/2012 7:41 PM
File folder

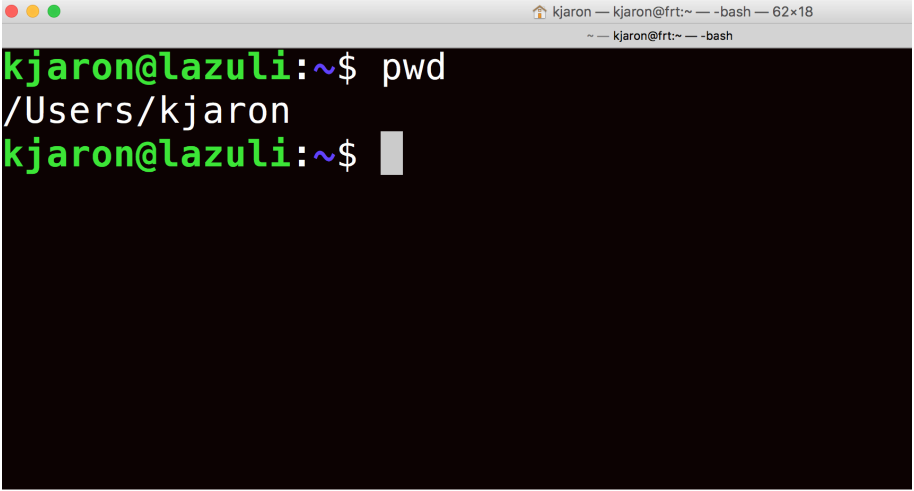


```
kjaron@frt:~$
```

user
computer
location
$ for user
# for administrator
space for a command

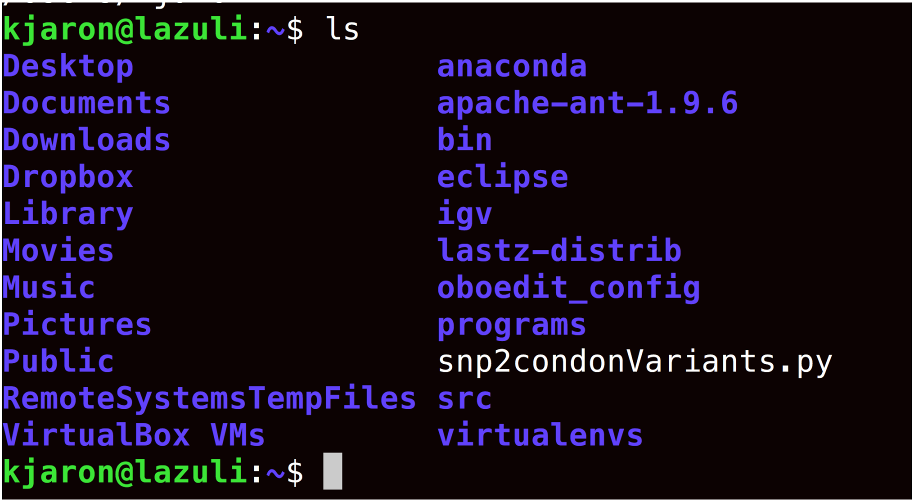
```
### relative paths
.     # this directory
..    # parent directory
~     # my home directory
### absolute path
/     # root directory
```

bash ≈ explorer (print working directory)



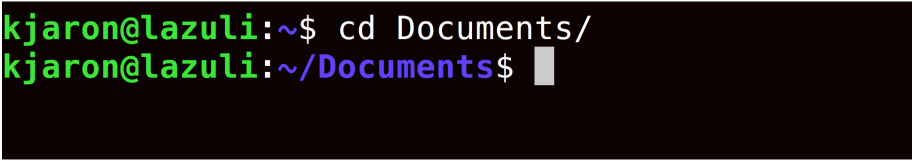
```
kjaron@lazuli:~$ pwd
/Users/kjaron
kjaron@lazuli:~$ 
```

bash ≈ explorer (list segments)



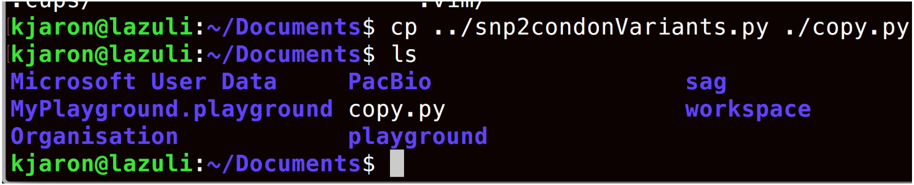
```
kjaron@lazuli:~$ ls
Desktop                anaconda
Documents              apache-ant-1.9.6
Downloads              bin
Dropbox                eclipse
Library                igv
Movies                 lastz-distrib
Music                  oboedit_config
Pictures               programs
Public                 snp2condonVariants.py
RemoteSystemsTempFiles src
VirtualBox VMs         virtualenvs
kjaron@lazuli:~$ █
```

bash ≈ explorer (change directory)



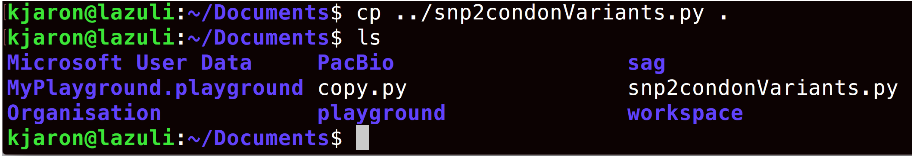
```
kjaron@lazuli:~$ cd Documents/
kjaron@lazuli:~/Documents$ 
```

bash ≈ explorer (copy)



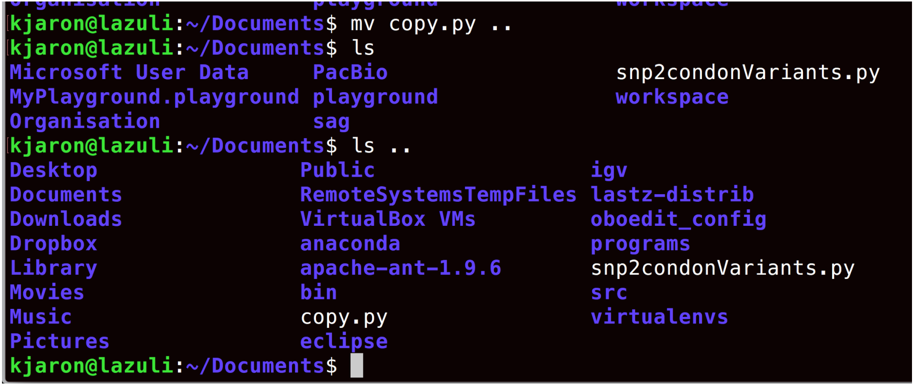
```
.cups/                          .vim/
[kjaron@lazuli:~/Documents$ cp ../snp2condonVariants.py ./copy.py
[kjaron@lazuli:~/Documents$ ls
Microsoft User Data      PacBio                   sag
MyPlayground.playground copy.py                  workspace
Organisation             playground
kjaron@lazuli:~/Documents$ 
```

bash ≈ explorer (copy)



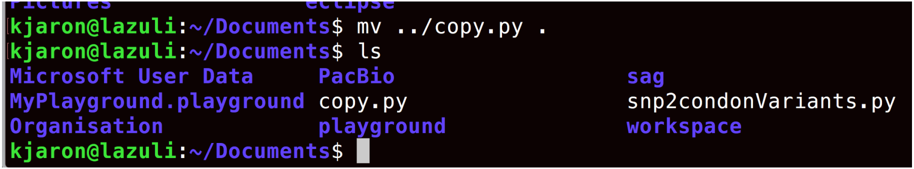
```
kjaron@lazuli:~/Documents$ cp ../snp2condonVariants.py .
kjaron@lazuli:~/Documents$ ls
Microsoft User Data      PacBio                  sag
MyPlayground.playground  copy.py                 snp2condonVariants.py
Organisation             playground              workspace
kjaron@lazuli:~/Documents$ 
```

bash ≈ explorer (move)



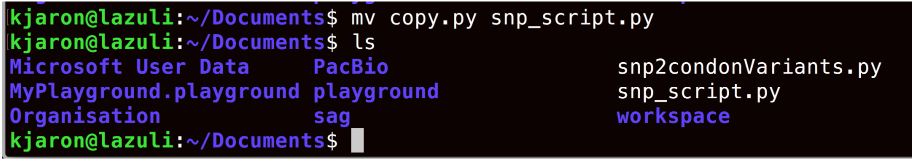
```
Organisation            playground              workspace
kjaron@lazuli:~/Documents$ mv copy.py ..
kjaron@lazuli:~/Documents$ ls
Microsoft User Data     PacBio                  snp2condonVariants.py
MyPlayground.playground playground              workspace
Organisation            sag
kjaron@lazuli:~/Documents$ ls ..
Desktop                Public                 igv
Documents              RemoteSystemsTempFiles lastz-distrib
Downloads              VirtualBox VMs         oboedit_config
Dropbox                anaconda               programs
Library                apache-ant-1.9.6       snp2condonVariants.py
Movies                 bin                    src
Music                  copy.py                virtualenvs
Pictures               eclipse
kjaron@lazuli:~/Documents$ 
```

bash ≈ explorer (move)



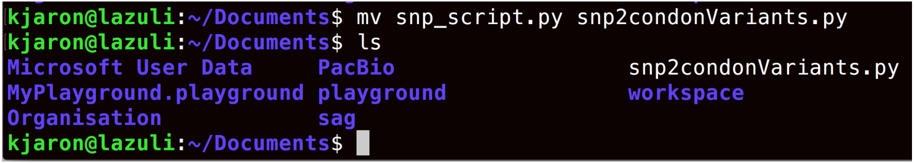
```
Pictures               eclipse
kjaron@lazuli:~/Documents$ mv ../copy.py .
kjaron@lazuli:~/Documents$ ls
Microsoft User Data     PacBio                  sag
MyPlayground.playground copy.py                 snp2condonVariants.py
Organisation            playground              workspace
kjaron@lazuli:~/Documents$ 
```

bash ≈ explorer (rename)



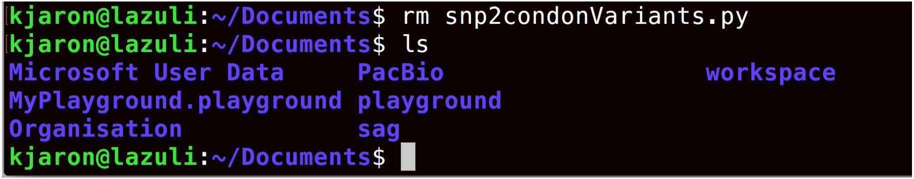
```
kjaron@lazuli:~/Documents$ mv copy.py snp_script.py
kjaron@lazuli:~/Documents$ ls
Microsoft User Data      PacBio                  snp2condonVariants.py
MyPlayground.playground  playground              snp_script.py
Organisation             sag                     workspace
kjaron@lazuli:~/Documents$ 
```

bash ≈ explorer (rename to existing file)



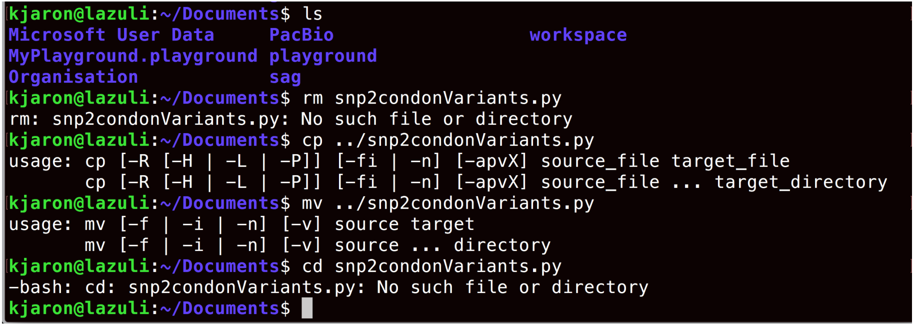
```
kjaron@lazuli:~/Documents$ mv snp_script.py snp2condonVariants.py
kjaron@lazuli:~/Documents$ ls
Microsoft User Data     PacBio                      snp2condonVariants.py
MyPlayground.playground playground                  workspace
Organisation            sag
kjaron@lazuli:~/Documents$ 
```

bash ≈ explorer (remove)



```
kjaron@lazuli:~/Documents$ rm snp2condonVariants.py
kjaron@lazuli:~/Documents$ ls
Microsoft User Data     PacBio                  workspace
MyPlayground.playground playground
Organisation            sag
kjaron@lazuli:~/Documents$ 
```

bash makes readable errors



```
kjaron@lazuli:~/Documents$ ls
Microsoft User Data     PacBio                  workspace
MyPlayground.playground playground
Organisation            sag
kjaron@lazuli:~/Documents$ rm snp2condonVariants.py
rm: snp2condonVariants.py: No such file or directory
kjaron@lazuli:~/Documents$ cp ../snp2condonVariants.py
usage: cp [-R [-H | -L | -P]] [-fi | -n] [-apvX] source_file target_file
       cp [-R [-H | -L | -P]] [-fi | -n] [-apvX] source_file ... target_directory
kjaron@lazuli:~/Documents$ mv ../snp2condonVariants.py
usage: mv [-f | -i | -n] [-v] source target
       mv [-f | -i | -n] [-v] source ... directory
kjaron@lazuli:~/Documents$ cd snp2condonVariants.py
-bash: cd: snp2condonVariants.py: No such file or directory
kjaron@lazuli:~/Documents$ 
```

```
command -<parameters> <arguments>
```

Examples:

```
ls -lah  #list long, all, human readable
ls -la ..  #list in parent directory
cp -r <what_dir> <were> # recursive
rm -rf <what dir>       # -||- , force
# careful with this one...

man <command>
```

Special characters *, ?, [], escape character \

```
$cd /
$echo b
b

$echo b*
bin boot

$echo b\*
b*

$echo B*
B*
```

OK, try it!

```
# Manual
man <command>


# Auto Completion by <tab>
cd /<tab><tab>    # lists all in root
cd ˜/k<tab><tab> # lists all in home
```

MyUnil → Sequence a Genome II → 1_unix_local.pdf

```
# Manual
man <command>

# Auto Completion by <tab>
cd /<tab><tab>   # lists all in root
cd ~/<tab><tab> # lists all in home

# Command history
<arrow_up>       # last excuted command
<Ctrl+R>         # full-text search
```

What about that toolbox?

How to check a file??

Notepad?

not well suited for big files

Office?

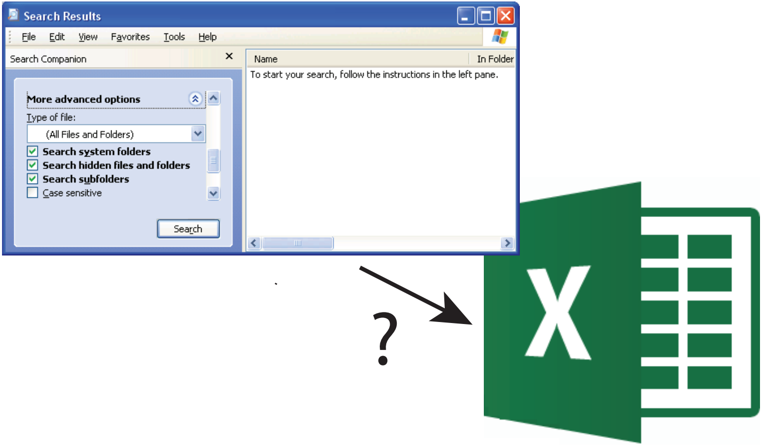
Use at least read-only mode

Look at a plain-text file

```
command <text_file>

echo
head / tail
cat   # catenate
less  # > more; text reader

wc    # word count
tr    # transform
grep  # global regular expression print
```

IO streams - a way how to build pipes!

```
command <text_file>
           == cat <text_file> | command
echo
tail / head
cat    # catenate
less   # > more

wc     # word count
tr     # transform
grep   # global regular expression print
```



Type of file: (All Files and Folders)

☑ Search system folders

☑ Search hidden files and folders

☑ Search subfolders

□ Case sensitive

To start your search, follow the instructions in the left pane.

IO streams!

```
ls | wc -w
```

Streams in the service of genomics!

```
grep ">" seq.fasta | wc -l
```

input stream (file)

```
grep -v ">" seq.fasta | \
  tr -d ’\n’ | wc -c
```

```
>seq1
CGATCGTCGTAGCTACGAT
>seq2
ACCGATCAAACCGTCGTAA
```

grep -v ">"

```
grep -v ">" seq.fasta | \
  tr -d "\n" | wc -c
```

```
CGATCGTCGTAGCTACGAT
ACCGATCAAACCGTCGTAA
```

tr -d "\n"

```
grep -v ">" seq.fasta | \
  tr -d "\n" | wc -c
```

```
CGATCGTCGTAGCTACGATACCGATCAAACCGTCGTAA
```

wc -c

```
grep -v ">" seq.fasta | \
  tr -d "\n" | wc -c
```

Stream redirection overview

```
grep ">" seq.fasta              # print
grep ">" seq.fasta | wc         # pipe
grep ">" seq.fasta > file       # write
grep ">" seq.fasta >> file # append
```

Your turn

```
command <parameters> <text_file>

echo
tail / head
cat   # catenate
less  # text reader

wc    # word count
tr    # transform
grep  # global regular expression print
```