

# Useful Commands for Wrangling Text Files

## Sorting your views line-by-line

---

`sort` - Sort lines in a file

```
$ sort pets.txt
cat
cat
cat
cat
cat
cat
dog
dog
```

```
$ sort -r pets.txt
dog
dog
cat
cat
cat
cat
cat
cat
```

## Excluding repeated lines from views

---

`uniq` (“unique”) – Filter out repeated lines

```
$ cat pets.txt
cat
cat
cat
cat
dog
dog
cat
```

```
$ uniq pets.txt
cat
dog
cat
```

## Excluding repeated lines from views

---

`uniq` (“unique”) – Filter out repeated lines

```
$ cat pets.txt
cat
cat
cat
cat
dog
dog
cat
```

```
$ uniq pets.txt
cat
dog
cat
```

## Extracting lines matching a pattern

---

`grep` (“global regular expression print”) – Return lines in file matching pattern

```
$ cat people.txt
Alan Turing
Bjarne Stroustrup
Charles Babbage
Dennis Ritchie
Erwin Schrodinger
Fred Hoyle
Guido Rossum
Henri Poincare
Ivan Pavlov
John Neumann
Ken Thompson
```

```
$ grep ch people.txt
Dennis Ritchie
Erwin Schrodinger

$ grep -i ch people.txt
Charles Babbage
Dennis Ritchie
Erwin Schrodinger
```

## Extracting slices from lines

`cut` – Extracts a section from each line

```
$ cat people.txt
```

```
Alan Turing  
Bjarne Stroustrup  
Charles Babbage  
Dennis Ritchie  
Erwin Schrodinger  
Fred Hoyle  
Guido Rossum  
Henri Poincare  
Ivan Pavlov  
John Neumann  
Ken Thompson
```

```
$ cut -c 2-9 people.txt
```

```
lan Turi  
jarne St  
harles B  
ennis Ri  
rwin Sch  
red Hoyle  
uido Ros  
enri Poi  
van Pavl  
ohn Neum  
en Thomp
```

## Extracting fields from lines

`cut` – Extract a field from each line

```
$ cat people.txt
```

```
Alan Turing  
Bjarne Stroustrup  
Charles Babbage  
Dennis Ritchie  
Erwin Schrodinger  
Fred Hoyle  
Guido Rossum  
Henri Poincare  
Ivan Pavlov  
John Neumann  
Ken Thompson
```

```
$ cut -d ' ' -f2 people.txt
```

```
Turing  
Stroustrup  
Babbage  
Ritchie  
Schrodinger  
Hoyle  
Rossum  
Poincare  
Pavlov  
Neumann  
Thompson
```

- Let's look at a more practical example for using `cut` with text files. Suppose you want to extract only the last name of each person on your list. You know that each line of your list is made up of two fields: the person's first and last names. A space separates these fields. Here, you're cutting to extract the second field from each line. Using the "minus d quote space quote" option, you can specify that the field delimiter, or the character that indicates the

break between fields, is a space. Then, use the “minus f two” option to return the second field from each line. This will return the last name of each person from the list.

## Merging lines from multiple files

---

paste – Merge lines from different files

```
$ cat first.txt
Alan
Bjarne
Charles
Dennis
Erwin
Fred
Guido
Henri
Ivan
John
Ken
```

## Merging lines from multiple files

---

paste – Merge lines from different files

```
$ cat last.txt
Turing
Stroustrup
Babbage
RSchrodinger
Hoyle
itchie
Rossum
Poincare
Pavlov
Neumann
Thompson
```

## Merging lines from multiple files

---

`paste` – Merge lines from different files

```
$ cat yob.txt
1912
1950
1791
1941
1887
1915
1956
1854
1849
1903
1943
```

## Merging lines from multiple files

---

`paste` – Merge lines from different files

```
$ paste first.txt last.txt yob.txt
Alan   Turing    1912
Bjarne Stroustrup 1950
Charles Babbage  1791
Dennis Ritchie   1941
Erwin  Schrodinger 1887
Fred   Hoyle     1915
Guido  Rossum      1956
Henri  Poincare   1854
Ivan   Pavlov     1849
John   Neumann    1903
Ken    Thompson   1943
```

# Merging lines from multiple files

---

paste – Merge lines from different files

```
$ paste -d "," first.txt last.txt yob.txt
Alan,Turing,1912
Bjarne,Stroustrup,1950
Charles,Babbage,1791
Dennis,Ritchie,1941
Erwin,Schrodinger,1887
Fred,Hoyle,1915
Guido,Rossum,1956
Henri,Poincare,1854
Ivan,Pavlov,1849
John,Neumann,1903
Ken,Thompson,1943
```

## Recap

---

In this video you learned how to:

- View the lines of a file alpha-numerically using `sort`
- Remove repeated lines from your view with `uniq`
- Get the lines of a file that match your desired criteria using `grep`
- Extract slices and fields from lines with `cut`