Using Grep

The grep (global regular expression print) command is used to search for patterns in a text file. The output of this command is the lines that match our search pattern. We use this command by saying:

grep (option) pattern (file name)

We can look at some examples below.

1. Print all lines containing the string Lane

grep Lane GrepLab

This command prints all the lines containing Lane. It works by printing all the lines that contain the word Lane in them, note this will only print the lines that match exactly and will ignore other possible variations, for example: lane, would not print as a result because it doesn't match our current search pattern.

```
[victoria@localhost ~]$ grep Lane GrepLab
Huckleberry Finn:238-923-7366:95 Latham Lane, Easton, PA 83755:11/12/56:20300
Betty Boop:245-836-8357:635 Cutesy Lane, Hollywood, CA 91464:6/23/23:14500
Lizzie Bennett:674-843-1385:20 Parak Lane, Duluth, MN 23850:4/12/23:780900
Nancy Drew:674-843-1385:20 Parak Lane, Duluth, MN 23850:4/12/23:780900
Ephram Hardy:293-259-5395:235 CarltonLane, Joliet, IL 73858:8/12/20:56700
Dorothy Gale:923-835-8745:23 Wimp Lane, Kensington, DL 38758:8/31/69:126000
Molly Weasley:387-827-1095:13 Uno Lane, Ashville, NC 23556:7/1/29:57000
[victoria@localhost ~]$
```

2. Print all lines where the person's first name starts with H

grep ^H GrepLab

^ this regex symbol matches a string that starts with.

By specifying ^H, we are telling grep to print all the lines in the file that start with 'H', this command is case sensitive, so it won't print lines that start with a 'h' unless we specify that. We can print all the lines that contain names starting with the letter H with this command because the names are at the beginning of the lines.

```
[victoria@localhost ~]$ grep ^H GrepLab
Huckleberry Finn:238-923-7366:95 Latham Lane, Easton, PA 83755:11/12/56:20300
Holly Golightly:397-857-2735:74 Pine Street, Dearborn, MI 23874:3/28/45:245700
Hester Prynne:408-253-3122:123 Park St., San Jose, CA 04086:7/25/53:85100
Hemione Granger:408-456-1234:4 Harvard Square, Boston, MA 02133:4/22/62:52600
[victoria@localhost ~]$
```

3. Print all lines ending in three zeros (000)

grep 000$ GrepLab

$ this regex symbol matches the end of the line.

This command prints lines that end on three zeros. Using the regex symbol **$** after the three zeros tells grep we want to print only the lines that specifically end with three zeros. **$** is useful for finding patterns at the end of the lines in a file.

```
[victoria@localhost ~]$ grep 000$ GrepLab
Meg Murry:834-938-8376:23445 Aster Ave., Allentown, NJ 83745:12/1/38:45000
Minerva McGonagall:408-233-8971:45 Rose Terrace, San Francisco, CA 92303:2/3/36:25000
Dorothy Gale:923-835-8745:23 Wimp Lane, Kensington, DL 38758:8/31/69:126000
Molly Weasley:387-827-1095:13 Uno Lane, Ashville, NC 23556:7/1/29:57000
[victoria@localhost ~]$
```

4. Print all lines that don't contain 408
   grep -v 408 GrepLab
   -v is an option in grep, it tells grep to print out the lines that don't match the given pattern. It won't print any lines containing '408' anywhere in the line.

```
Huckleberry Finn:238-923-7366:95 Latham Lane, Easton, PA 83755:11/12/56:20300
Betty Boop:245-836-8357:635 Cutesy Lane, Hollywood, CA 91464:6/23/23:14500
Dorian Gray:385-375-8395:3567 Populus Place, Caldwell, NJ 23875:6/18/60:23400
Holly Golightly:397-857-2735:74 Pine Street, Dearborn, MI 23874:3/28/45:245700
Ebenezer Scrooge:548-834-2348:583 Laurel Ave., Kingsville, TX 83745:10/1/35:58900
Westley Pirate:284-758-2857:23 Edgecliff Place, Lincoln , NB 92743:7/25/53:85100
Westley Pirate:284-758-2867:23 Edgecliff Place, Lincoln, NB 92743:11/3/35:58200
Westley Pirate:284-758-2867:23 Edgecliff Place, Lincoln, NB 92743:11/3/35:58200
Lizzie Bennett:674-843-1385:20 Parak Lane, Duluth, MN 23850:4/12/23:780900
Nancy Drew:674-843-1385:20 Parak Lane, Duluth, MN 23850:4/12/23:780900
Jo March:327-832-5728:3465 Mirlo Street, Peabody, MA 34756:10/2/65:35200
Victor Frankenstein:835-365-1204:454 Easy Street, Decatur, IL 75732:2/28/53:123500
Ephram Hardy:293-259-5395:235 CarltonLane, Joliet, IL 73858:8/12/20:56700
Meg Murry:834-938-8376:23445 Aster Ave., Allentown, NJ 83745:12/1/38:45000
Lucy Pevensie:385-573-8326:832 Ponce Drive, Gzary, IN 83756:12/1/46:268500
Mary Poppins:846-836-2837:6937 Ware Road, Milton, PA 93756:9/21/46:43500
Sir Lancelot:837-835-8257:474 Camelot Boulevard, Bath, WY 28356:5/13/69:24500
Zippy Pinhead:834-823-8319:2356 Bizarro Ave., Farmount, IL 84357:1/1/67:89500
Dorothy Gale:923-835-8745:23 Wimp Lane, Kensington, DL 38758:8/31/69:126000
Popeye Sailor:156-454-3322:945 Bluto Street, NotHere, USA 29358:3/19/35:22350
Luna Lovegood:385-898-8357:38 Fife Way, Abilene, TX 39673:1/5/58:95600
Molly Weasley:387-827-1095:13 Uno Lane, Ashville, NC 23556:7/1/29:57000
Arya Stark:438-910-7449:8235 Maple Street, Wilmington, VM 29085:9/23/63:68900
[victoria@localhost ~]$
```

5. Print all lines where birthdays are in the year 1935 (be careful of the

date format! it's MM/DD/YY)

grep /35 GrepLab

In this document the only instances for "/35" are for the birth years, it prints the lines including birthdays in 1935.

```
[victoria@localhost ~]$ grep /35 GrepLab
Ebenezer Scrooge:548-834-2348:583 Laurel Ave., Kingsville, TX 83745:10/1/35:58900
Westley Pirate:284-758-2867:23 Edgecliff Place, Lincoln, NB 92743:11/3/35:58200
Westley Pirate:284-758-2867:23 Edgecliff Place, Lincoln, NB 92743:11/3/35:58200
Popeye Sailor:156-454-3322:945 Bluto Street, NotHere, USA 29358:3/19/35:22350
[victoria@localhost ~]$
```

6. Print all lines where the phone number is in an area code that starts

with an 8

grep 8[0-9][0-9][-][0-9][0-9][0-9][-][0-9][0-9][0-9][0-9]

[0-9] is a regex that matches any numbers between 0-9

This command uses a regex pattern to search and print lines containing phone numbers with area codes that start with 8.

'8' at the beginning of the pattern matches the literal number at the beginning of the area code.

'[0-9][0-9]' matches any 2 numbers after the 8 in the area code.

'[-]' matches the hyphen separating the area code from the rest of the number.

'[0-9][0-9][0-9]' matches any three numbers (0-9) for the next three digits of the phone number.

'[-]' matches the hyphen separating the first three numbers from the last four of the phone number.

'[0-9][0-9][0-9][0-9]' matches any four numbers for the last four digits of the phone number.

Using this regex we ensure that we are printing any line containing a phone number with an area code that start with the number 8 regardless of what the numbers are following.

```
[victoria@localhost ~]$ grep 8[0-9][0-9][-][0-9][0-9][0-9][-][0-9][0-9][0-9][0-9] GrepLab
Victor Frankenstein:835-365-1284:454 Easy Street, Decatur, IL 75732:2/28/53:123500
Meg Murry:834-938-8376:23445 Aster Ave., Allentown, NJ 83745:12/1/38:45000
Mary Poppins:846-836-2837:6937 Ware Road, Milton, PA 93756:9/21/46:43500
Sir Lancelot:837-835-8257:474 Camelot Boulevard, Bath, WY 28356:5/13/69:24500
Zippy Pinhead:834-823-8319:2356 Bizarro Ave., Farmount, IL 84357:1/1/67:89500
[victoria@localhost ~]$
```

7. Print all lines containing an uppercase letter, followed by 4 lowercase

letters, a space and one uppercase letter.

grep '[A-Z][a-z]{4} [A-Z]' GrepLab

[A-Z] matches any single capital letter.

[a-z]{4} matches any lower case letter the {4} indicates we want this done four times, meaning we are looking for the occurrence of 4 consecutive lowercase letters (in this particular pattern, specifically after an uppercase letter).

Notice there is a space after the closing curly bracket and the opening square bracket, this indicates we are looking for an empty space that matches our pattern.

[A-Z] matches any single capital letter.

This command uses regex to search and print any occurrence of an uppercase letter immediately followed by 4 lowercase letters, a space and one uppercase letter. It does this regardless of where in the line it finds the matching pattern.

```
victoria@ubuntuserver:~$ grep '[A-Z][a-z]{4} [A-Z]' GrepLab
victoria@ubuntuserver:~$ grep -E '[A-Z][a-z]{4} [A-Z]' GrepLab
Betty Boop:245-836-8357:635 Cutesy Lane, Hollywood, CA 91464:6/23/23:14500
Holly Golightly:397-857-2735:74 Pine Street, Dearborn, MI 23874:3/28/45:245700
Lizzie Bennett:674-843-1385:20 Parak Lane, Duluth, MN 23850:4/12/23:780900
Nancy Drew:674-843-1385:20 Parak Lane, Duluth, MN 23850:4/12/23:780900
Jo March:327-832-5728:3465 Mirlo Street, Peabody, MA 34756:10/2/65:35200
Meg Murry:834-938-8376:23445 Aster Ave., Allentown, NJ 83745:12/1/38:45000
Lucy Pevensie:385-573-8326:832 Ponce Drive, Gzary, IN 83756:12/1/46:268500
Zippy Pinhead:834-823-8319:2356 Bizarro Ave., Farmount, IL 84357:1/1/67:89500
Popeye Sailor:156-454-3322:945 Bluto Street, NotHere, USA 29358:3/19/35:22350
Daenerys Targaryen:408-724-0140:1222 Oxbow Court, Sunnyvale, CA 94087:5/19/66:34200
Molly Weasley:387-827-1095:13 Uno Lane, Ashville, NC 23556:7/1/29:57000
Arya Stark:438-910-7449:8235 Maple Street, Wilmington, VM 29085:9/23/63:68900
```

8. Print lines where the address begins with a two or three digit number

(so this would be 12 main st or 123 main street but not 1234 main

street).

grep -E ':[0-9]{2,3} [A-Za-z]+ ' GrepLab

-E this option is for extended regular expressions, this option allows for more flexible pattern matching, making it easier to write and understand regex patterns.

:[0-9]{2,3}this combination matches any number containing 2 or 3 digits happening after the column (:). This {2,3} implies we are looking for any 2 or 3 digit number.

Notice the space after the closing curly bracket and the square bracket, this means we are looking for a space after the numbers.

[A-Za-z] matches any uppercase or lowercase letter, and the + after the square bracket means that our letter pattern should happen one or more times.

```
victoria@ubuntuserver:~$ grep -E ':[0-9]{2,3} [A-Za-z]+' GrepLab
Huckleberry Finn:238-923-7366:95 Latham Lane, Easton, PA 83755:11/12/56:20300
Betty Boop:245-836-8357:635 Cutesy Lane, Hollywood, CA 91464:6/23/23:14500
Holly Golightly:397-857-2735:74 Pine Street, Dearborn, MI 23874:3/28/45:245700
Ebenezer Scrooge:548-834-2348:583 Laurel Ave., Kingsville, TX 83745:10/1/35:58900
Hester Prynne:408-253-3122:123 Park St., San Jose, CA 04086:7/25/53:85100
Westley Pirate:284-758-2857:23 Edgecliff Place, Lincoln , NB 92743:7/25/53:85100
Westley Pirate:284-758-2867:23 Edgecliff Place, Lincoln, NB 92743:11/3/35:58200
Westley Pirate:284-758-2867:23 Edgecliff Place, Lincoln, NB 92743:11/3/35:58200
Lizzie Bennett:674-843-1385:20 Parak Lane, Duluth, MN 23850:4/12/23:780900
Nancy Drew:674-843-1385:20 Parak Lane, Duluth, MN 23850:4/12/23:780900
Victor Frankenstein:835-365-1284:454 Easy Street, Decatur, IL 75732:2/28/53:123500
Ephram Hardy:293-259-5395:235 CarltonLane, Joliet, IL 73858:8/12/20:56700
Lucy Pevensie:385-573-8326:832 Ponce Drive, Gzary, IN 83756:12/1/46:268500
Sir Lancelot:837-835-8257:474 Camelot Boulevard, Bath, WY 28356:5/13/69:24500
Minerva McGonagall:408-233-8971:45 Rose Terrace, San Francisco, CA 92303:2/3/36:25000
Dorothy Gale:923-835-8745:23 Wimp Lane, Kensington, DL 38758:8/31/69:126000
Popeye Sailor:156-454-3322:945 Bluto Street, NotHere, USA 29358:3/19/35:22350
Luna Lovegood:385-898-8357:38 Fife Way, Abilene, TX 39673:1/5/58:95600
Molly Weasley:387-827-1095:13 Uno Lane, Ashville, NC 23556:7/1/29:57000
```

9. Print lines where the person lives in Mass or Illinois

grep "MA\|IL" GrepLab

\| is an operator and it works as saying OR, this allows for multiple patterns to be matched.

In this document, states are represented by their state abbreviations which are usually a combination of two capital letters. This command works because it prints the lines containing the requested states there for also printing the person that lives in that state.

```
victoria@ubuntuserver:~$ grep "MA\|IL" GrepLab
Jo March:327-832-5728:3465 Mirlo Street, Peabody, MA 34756:10/2/65:35200
Victor Frankenstein:835-365-1284:454 Easy Street, Decatur, IL 75732:2/28/53:123500
Ephram Hardy:293-259-5395:235 CarltonLane, Joliet, IL 73858:8/12/20:56700
Hemione Granger:408-456-1234:4 Harvard Square, Boston, MA 02133:4/22/62:52600
Zippy Pinhead:834-823-8319:2356 Bizarro Ave., Farmount, IL 84357:1/1/67:89500
```

10. Print lines containing the addresses that aren't on a street (You might

see St as shorthand for street)

grep -viE '\b(st\.?|street)\b' GrepLab

this command uses a combination of grep options and regular expressions. In this case we are grouping different options together (-viE) and we are going to break them down below to see how they work

 -v this option matches lines that don't contain the specified pattern.

- i this option makes the command insensitive to case.

-E extended regular expression that lets us have more flexibility with regex.

\b (we have this at the beginning and end of our regex pattern) is a word boundary, it ensures the match happens at the beginning or end of a word.

(st\.?|street) this pattern matches either "st." or "street", since we made it case insensitive with the i option this also will match instances where there are capital letters.

| this operator means or so it states we are looking for st, or street.

The combination of operators and regex used in this command prints lines that do not contain st. or street also ignoring variations with capital letters.

```
Westley Pirate:284-758-2867:23 Edgecliff Place, Lincoln, NB 92743:11/3/35:58200
Westley Pirate:284-758-2867:23 Edgecliff Place, Lincoln, NB 92743:11/3/35:58200
Lizzie Bennett:674-843-1385:20 Parak Lane, Duluth, MN 23850:4/12/23:780900
Nancy Drew:674-843-1385:20 Parak Lane, Duluth, MN 23850:4/12/23:780900


Ephram Hardy:293-259-5395:235 CarltonLane, Joliet, IL 73858:8/12/20:56700
Meg Murry:834-938-8376:23445 Aster Ave., Allentown, NJ 83745:12/1/38:45000
Lucy Pevensie:385-573-8326:832 Ponce Drive, Gzary, IN 83756:12/1/46:268500
Hemione Granger:408-456-1234:4 Harvard Square, Boston, MA 02133:4/22/62:52600
Mary Poppins:846-836-2837:6937 Ware Road, Milton, PA 93756:9/21/46:43500
Sir Lancelot:837-835-8257:474 Camelot Boulevard, Bath, WY 28356:5/13/69:24500
Minerva McGonagall:408-233-8971:45 Rose Terrace, San Francisco, CA 92303:2/3/36:25000
Zippy Pinhead:834-823-8319:2356 Bizarro Ave., Farmount, IL 84357:1/1/67:89500
Dorothy Gale:923-835-8745:23 Wimp Lane, Kensington, DL 38758:8/31/69:126000

Luna Lovegood:385-898-8357:38 Fife Way, Abilene, TX 39673:1/5/58:95600
Daenerys Targaryen:408-724-0140:1222 Oxbow Court, Sunnyvale, CA 94087:5/19/66:34200
Molly Weasley:387-827-1095:13 Uno Lane, Ashville, NC 23556:7/1/29:57000
victoria@ubuntuserver:~$
```

References

https://www.guru99.com/linux-regular-expressions.html

https://www.gnu.org/software/grep/manual/grep.html

https://www.youtube.com/watch?v=aZGUBSQDAdI&t=9s&ab_channel=KrisOcchipinti

https://www.thegeekstuff.com/2011/01/advanced-regular-expressions-in-grep-command-with-10-examples-%E2%80%93-part-ii/