CS5354 UNIX TOOL PROGRAMMING

grep

- A. Sawarkar

Regular Expressions grep and egrep

Previously

- Basic UNIX Commands
 - Files: rm, cp, mv, ls, ln
 - Processes: ps, kill
- Unix Filters
 - cat, head, tail, tee, wc
 - cut, paste
 - find
 - sort, uniq
 - tr

Today

- Regular Expressions
 - Allow you to search for text in files
 - grep command
- Stream manipulation:
 - -sed

Regular Expressions

What Is a Regular Expression?

- A regular expression (regex) describes a set of possible input strings.
- Regular expressions is a fundamental concept in Computer Science called finite automata theory
- Regular expressions of Unix
 - vi, ed, sed, and emacs
 - awk, tcl, perl and Python
 - grep, egrep, fgrep
 - compilers

Regular Expressions

- The simplest regular expressions are a string of literal characters to match.
- The string matches the regular expression if it contains the substring.





| match

UNIX Tools sucks.

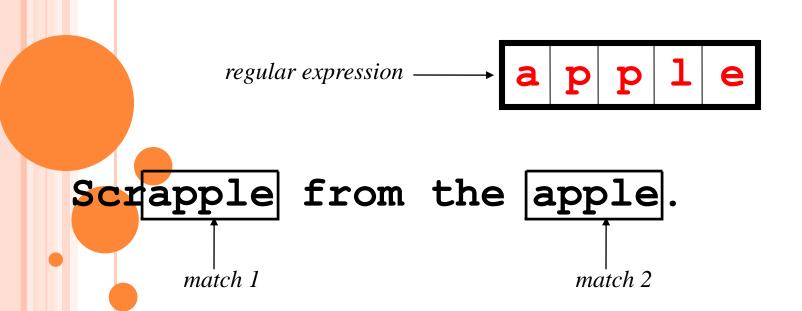
match

UNIX Tools is okay.

no match

Regular Expressions

 A regular expression can match a string in more than one place.



grep

- **grep** comes from the **ed** (Unix text editor) search command "global regular expression print" or **g/re/p**
 - This was such a useful command that it was written as a standalone utility
- There are two other variants, *egrep* and *fgrep* that comprise the *grep* family
 - grep is the answer to the moments where you know you want the file that contains a specific phrase but you can't remember its name

Family Differences

- grep uses regular expressions for pattern matching
- **fgrep** file grep, does not use regular expressions, only matches fixed strings but can get search strings from a file
- egrep extended grep, uses a more powerful set of regular expressions but does not support backreferencing, generally the fastest member of the grep family
- **agrep** approximate grep; not standard

Grep Syntax

grep: Searching for a pattern

-scan its input for pattern and display lines containing the pattern.

Syntax

grep [options] pattern filename(s)
grep [-hilnv] [-e expression] [filename]

e.g \$ grep "sales" emp.lst

[display lines contain "sales" from file emp.lst.

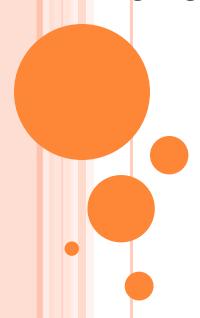
\$ grep "sales" emp.lst

```
adhoc@adhoc: ~/Desktop
File Edit View Search Terminal Help
adhoc@adhoc:~/Desktop$ cat emp.lst
       a.k.shukla
2233
                                       sales
                            q.m.
                                                     12/12/52
                                                                6000
9876
       jai sharma
                            director
                                       production
                                                     12/03/50
                                                                7000
       sumit chakrobarty
5678
                            d.g.m.
                                       marketing
                                                    19/04/43
                                                                6000
2365
       barun sengupta
                            director
                                       personnel
                                                    11/05/47
                                                                7800
5423
       n.k.gupta
                           chairman I
                                       admin
                                                     30/08/56
                                                                5400
       chanchal singhvi
                                       sales
1006
                           director I
                                                    03/09/38
                                                                6700
6213
       karuna ganguly
                                       accounts
                                                    05/06/62
                                                                6300
                           g.m.
1265
      s.n. dasqupta
                                       sales
                                                     12/09/63
                                                                5600
                           manager
4290
       jayant choudhury
                            executive|
                                       production
                                                    07/09/50
                                                                6000
       anil aggarwal
                                       sales
2476
                           manager
                                                    01/05/59
                                                                5000
                           director
                                       marketing
6521
       lalit chowdury
                                                     26/09/45
                                                                8200
       shyam saksena
                                       accounts
3212
                            d.q.m.
                                                    12/12/55
                                                                6000
       sudhir agarwal
                            executive
                                       personnel
3564
                                                    06/07/47
                                                                7500
2345
       j. b. sexena
                                       marketing
                                                    12/03/45
                                                                8000
                            q.m.
0110
       v.k.agrawal
                                       marketing
                                                     31/12/40
                                                                9000
                            q.m.
adhoc@adhoc:~/Desktop$ grep
                              "sales"
                                      emp.lst
       a.k.shukla
2233
                                       sales
                                                    12/12/52
                                                                6000
                           q.m.
       chanchal singhvi
1006
                            director
                                       sales
                                                    03/09/38
                                                                6700
                                                    12/09/63
1265
       s.n. dasgupta
                                       sales
                                                                5600
                            manager
                                      sales
       anil aggarwal
2476
                                                    01/05/59
                                                                5000
                           manager
adhoc@adhoc:~/Desktop$
```

Grep with multiple file

\$grep "director" emp1.lst emp2.lst
[Quoting (' ') is essential when the pattern contains multiple word .

e.g \$ grep 'jai sharma' emp.lst]



grep options:

-i	Ignore case for matching
-v	Doesn't display lines matching expression
-n	Display line numbers along with lines
- c	Display count of number of occurrence
-1	Display list of filename only
-e	expression
- x	Match pattern with entire line
-f	File take pattern from file one per line
-E	Treats patterns as an extended regular expr
-F	Match multiple fixed string

-i: Ignoring Case

\$ grep -i 'agarwal' emp.lst

```
adhoc@adhoc: ~/Desktop
 File Edit View Search Terminal Help
 adhoc@adhoc:~/Desktop$ cat emp.lst
©2233 ∣
        a.k.shukla
                                        sales
                                                     12/12/52
                                                                 6000
                            g.m.
 9876
        jai sharma
                            director I
                                        production
                                                     12/03/50
                                                                 7000
 5678
        sumit chakrobarty
                                        marketing
                            d.g.m.
                                                     19/04/43
                                                                 6000
 2365
        barun sengupta
                            director |
                                        personnel
                                                     11/05/47
                                                                 7800
 5423
        n.k.gupta
                             chairman I
                                        admin
                                                     30/08/56
                                                                 5400
        chanchal singhvi
 1006
                           | director |
                                        sales
                                                     03/09/38
                                                                 6700
 6213 I
        karuna ganguly
                                        accounts
                                                     05/06/62
                                                                 6300
                           q.m.
 1265
        s.n. dasgupta
                                        sales
                                                     12/09/63
                                                                 5600
                            manager
 4290
        jayant choudhury
                             executivel
                                        production
                                                     07/09/50
                                                                 6000
 2476 I
        anil aggarwal
                                        sales
                                                     01/05/59
                                                                 5000
                            manager
 6521 l
        lalit chowdury
                            director |
                                        marketing
                                                     26/09/45
                                                                 8200
 3212 I
        shyam saksena
                            d.q.m.
                                        accounts
                                                     12/12/55
                                                                 6000
3564
        sudhir Agarwal
                                        personnel
                             executive
                                                     06/07/47
                                                                 7500
 2345
        j. b. sexena
                                        marketing
                                                     12/03/45
                                                                 8000
                             q.m.
 0110
      | v.k.agrawal
                                        marketing
                                                     31/12/40
                                                                 9000
                            q.m.
 adhoc@adhoc:~/Desktop$ grep -i 'agarwal' emp.lst
        sudhir Agarwal
                           | executive| personnel
                                                     06/07/47
 adhoc@adhoc:~/Desktop$
```

-v: Deleting line (Inverse)

\$ grep -v 'director' emp.lst > otherlist [create]

```
adhoc@adhoc: ~/Desktop
 File Edit View Search Terminal Help
adhoc@adhoc:~/Desktop$ cat emp.lst
©2233
       a.k.shukla
                                       sales
                                                     12/12/52
                                                                 6000
                            q.m.
9876
     I jai sharma
                                       production
                                                     12/03/50
                            director
                                                                 7000
5678
       sumit chakrobarty
                                       marketing
                                                     19/04/43
                            d.q.m.
                                                                 6000
       barun sengupta
2365
                            director
                                        personnel
                                                     11/05/47
                                                                 7800
       n.k.qupta
5423
                            chairman
                                        admin
                                                     30/08/56
                                                                 5400
1006
       chanchal singhvi
                                        sales
                                                     03/09/38
                                                                6700
                            director
       karuna ganguly
6213
                                                     05/06/62
                                                                6300
                            g.m.
                                        accounts
       s.n. dasgupta
                                                     12/09/63
1265
                            manager
                                                                 5600
                                        sales
       jayant choudhury
4290
                            executive| production
                                                     07/09/50
                                                                 6000
       anil aggarwal
2476 I
                            manager
                                        sales
                                                     01/05/59
                                                                 5000
6521
       lalit chowdury
                            director
                                       marketing
                                                     26/09/45
                                                                8200
       shvam saksena
3212
                            d.g.m.
                                        accounts
                                                     12/12/55
                                                                 6000
       sudhir Agarwal
                                                     06/07/47
3564
                            executive| personnel
                                                                 7500
       j. b. sexena
                                       marketing
                                                     12/03/45
                                                                8000
                            q.m.
0110 | v.k.agrawal
                                       marketing
                                                     31/12/40
                            q.m.
                                                                 9000
adhoc@adhoc:~/Desktop$ grep -v 'director' emp.lst > otherlist
adhoc@adhoc:~/Desktop$ cat otherlist
2233 I
       a.k.shukla
                                        sales
                                                     12/12/52
                            q.m.
                                                                 6000
5678
       sumit chakrobarty |
                                       marketing
                                                                 6000
                            d.g.m.
                                                     19/04/43
       n.k.gupta
 5423
                            chairman
                                        admin
                                                     30/08/56
                                                                 5400
       karuna ganguly
                                                     05/06/62
6213
                            g.m.
                                        accounts
                                                                6300
1265
       s.n. dasgupta
                                       sales
                                                                 5600
                            manager
                                                     12/09/63
     | jayant choudhury
                            executive| production
 4290
                                                     07/09/50
                                                                6000
2476
       anil aggarwal
                                                     01/05/59
                                       sales
                                                                 5000
                            manager
       shvam saksena
3212
                            d.q.m.
                                        accounts
                                                     12/12/55
                                                                 6000
       sudhir Agarwal
                            executive| personnel
3564
                                                     06/07/47
                                                                7500
2345
       j. b. sexena
                                       marketing
                                                     12/03/45
                                                                 8000
                            q.m.
0110 | v.k.agrawal
                                       marketing
                            q.m.
                                                     31/12/40
                                                                 9000
adhoc@adhoc:~/Desktop$
```

-n: Displaying Line numbers

\$grep -n marketing' emp.lst

```
adhoc@adhoc: ~/Desktop
File Edit View Search Terminal Help
adhoc@adhoc:~/Desktop$ cat emp.lst
2233
       a.k.shukla
                                      sales
                                                                6000
                           q.m.
                                                    12/12/52
9876
      iai sharma
                           director
                                       production |
                                                               7000
                                                    12/03/50
5678
       sumit chakrobarty
                           d.g.m.
                                      marketing
                                                    19/04/43
                                                               6000
2365
       barun sengupta
                           director |
                                       personnel
                                                    11/05/47
                                                               7800
5423
      n.k.gupta
                           chairman
                                       admin
                                                    30/08/56
                                                               5400
       chanchal singhvi
1006
                           director
                                       sales
                                                    03/09/38
                                                               6700
6213
       karuna ganguly
                                       accounts
                                                    05/06/62
                                                               6300
                           g.m.
                                                    12/09/63
1265 |
       s.n. dasgupta
                                                               5600
                           manager
                                       sales
4290 | jayant choudhury
                                                    07/09/50
                                                               6000
                           executive| production |
2476
       anil aggarwal
                                       sales
                                                    01/05/59
                                                               5000
                           manager
6521 | lalit chowdury
                           director |
                                      marketing
                                                    26/09/45
                                                               8200
3212 I
       shyam saksena
                                                    12/12/55
                                                               6000
                           d.g.m.
                                       accounts
3564
       sudhir Agarwal
                           executive|
                                      personnel
                                                    06/07/47
                                                               7500
2345 I
      j. b. sexena
                                       marketing
                                                    12/03/45
                                                               8000
                           g.m.
0110 | v.k.agrawal
                                       marketing
                                                    31/12/40
                                                               9000
                           q.m.
adhoc@adhoc:~/Desktop$ grep -n 'marketing' emp.lst
3:5678 | sumit chakrobarty |
                                        marketing
                             d.g.m.
                                                     19/04/43
                                                                 6000
                                          marketing
11:6521 | lalit chowdury
                              director
                                                       26/09/45
                                                                  8200
                                          marketing
14:2345
         j. b. sexena
                              q.m.
                                                       12/03/45
                                                                   8000
15:0110 | v.k.agrawal
                                          marketing
                                                       31/12/40
                                                                   9000
                              g.m.
adhoc@adhoc:~/Desktop$
```

-c :counting line containg pattern

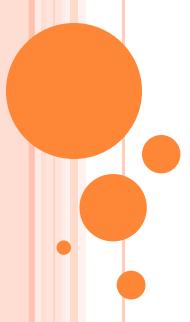
\$grep -c 'director' emp.lst

```
adhoc@adhoc: ~/Desktop
 File Edit View Search Terminal Help
adhoc@adhoc:~/Desktop$ cat emp.lst
2233
       a.k.shukla
                            q.m.
                                       sales
                                                     12/12/52
                                                                6000
9876
       jai sharma
                            director
                                       production
                                                     12/03/50
                                                                7000
5678
       sumit chakrobarty
                            d.g.m.
                                       marketing
                                                     19/04/43
                                                                6000
2365
       barun sengupta
                            director
                                       personnel
                                                     11/05/47
                                                                7800
       n.k.gupta
5423
                            chairman
                                       admin
                                                     30/08/56
                                                                5400
       chanchal singhvi
                            director
1006
                                       sales
                                                     03/09/38
                                                                6700
6213
       karuna ganguly
                                                     05/06/62
                                                                6300
                                       accounts
                            g.m.
1265
       s.n. dasqupta
                                       sales
                                                     12/09/63
                                                                5600
                            manager
       jayant choudhury
4290
                            executive| production
                                                     07/09/50
                                                                6000
       anil aggarwal
                                       sales
2476
                            manager
                                                     01/05/59
                                                                5000
       lalit chowdury
6521
                            director |
                                       marketing
                                                     26/09/45
                                                                8200
3212
       shyam saksena
                            d.g.m.
                                       accounts
                                                     12/12/55
                                                                6000
       sudhir Agarwal
                            executive|
                                       personnel
3564
                                                     06/07/47
                                                                7500
2345
       b. sexena
                                       marketing
                                                     12/03/45
                                                                8000
                            q.m.
0110
     | v.k.agrawal
                                       marketing
                            q.m.
                                                     31/12/40
                                                                9000
adhoc@adhoc:~/Desktop$ grep -c 'director' emp.lst
adhoc@adhoc:~/Desktop$
```

-c :counting line containg pattern More than one files

\$grep -c 'director' emp*.lst

```
View Search Terminal Help
adhoc@adhoc:~/Desktop$ cat emp.lst
       a.k.shukla
                            g.m.
                                        sales
                                                      12/12/52
                                                                  6000
9876
       jai sharma
                            director
                                        production
                                                      12/03/50
                                                                  7000
5678
       sumit chakrobarty
                                        marketing
                            d.q.m.
                                                      19/04/43
                                                                  6000
                                        personnel
2365
       barun sengupta
                            director
                                                      11/05/47
                                                                  7800
5423
       n.k.gupta
                            chairman
                                        admin
                                                      30/08/56
                                                                  5400
1006
       chanchal singhvi
                            director
                                        sales
                                                      03/09/38
                                                                  6700
6213
       karuna ganguly
                                        accounts
                                                      05/06/62
                                                                  6300
                            q.m.
1265
       s.n. dasgupta
                            manager
                                        sales
                                                      12/09/63
                                                                  5600
       jayant choudhury
                                        production
4290
                            executive
                                                      07/09/50
                                                                  6000
2476
       anil aggarwal
                            manager
                                        sales
                                                      01/05/59
                                                                  5000
6521
       lalit chowdury
                            director
                                        marketing
                                                      26/09/45
                                                                  8200
3212
       shvam saksena
                            d.g.m.
                                        accounts
                                                      12/12/55
                                                                  6000
3564
       sudhir Agarwal
                            executive|
                                        personnel
                                                      06/07/47
                                                                  7500
2345
       i. b. sexena
                                        marketing
                                                      12/03/45
                                                                  8000
       v.k.agrawal
                            q.m.
                                        marketing
                                                      31/12/40
                                                                  9000
adhoc@adhoc:~/Desktop$ cat
                            emp1.lst
       a.k.shukla
                                        sales
                                                      12/12/52
                                                                  6000
                            a.m.
9876
       jai sharma
                            director
                                        production
                                                      12/03/50
                                                                  7000
       sumit chakrobarty
                                        marketing
                            d.q.m.
                                                      19/04/43
                                                                  6000
       barun sengupta
                            director
                                        personnel
                                                      11/05/47
                                                                  7800
adhoc@adhoc:~/Desktop$ cat
                            emp2.lst
       a.k.shukla
2233
                            q.m.
                                        sales
                                                      12/12/52
                                                                  6000
9876
       jai sharma
                            director
                                        production
                                                      12/03/50
                                                                  7000
       sumit chakrobarty
                                        marketing
                            d.q.m.
                                                      19/04/43
                                                                  6000
2365
       barun sengupta
                            director
                                        personnel
                                                      11/05/47
                                                                  7800
5423
       n.k.gupta
                            chairman
                                        admin
                                                      30/08/56
                                                                  5400
1006
       chanchal singhvi
                            director
                                        sales
                                                      03/09/38
                                                                  6700
6213
       karuna ganguly
                                                      05/06/62
                                                                  6300
                                        accounts
1265
       s.n. dasgupta
                            manager
                                        sales
                                                      12/09/63
                                                                  5600
       javant choudhury
                            executive| production
                                                      07/09/50
                                                                  6000
adhoc@adhoc:~/Desktop$ grep -c 'director' emp*.lst
emp1.lst:2
emp2.lst:3
emp.lst:4
adhoc@adhoc:~/Desktop$
```



-l:Displaying filenames

\$grep -l 'manager' *.lst

```
adhoc@adhoc: ~/Desktop

File Edit View Search Terminal Help

adhoc@adhoc: ~/Desktop$ grep -l 'manager' *.lst

desigx.lst

emp2.lst

emp.lst
adhoc@adhoc: ~/Desktop$
```

-e: Matching multiple patterns

\$grep -e "Agarwal" -e "aggarwal" -e "agrawal" emp.lst

```
adhoc@adhoc: ~/Desktop
 File Edit View Search Terminal Help
adhoc@adhoc:~/Desktop$ cat emp.lst
 2233
        a.k.shukla
                            g.m.
                                        sales
                                                                 6000
                                                     12/12/52
       jai sharma
                            director
                                        production
                                                     12/03/50
©9876
                                                                7000
        sumit chakrobarty
 5678
                            d.a.m.
                                        marketing
                                                     19/04/43
                                                                6000
 2365
        barun sengupta
                            director
                                        personnel
                                                     11/05/47
                                                                7800
 5423
        n.k.gupta
                            chairman
                                        admin
                                                     30/08/56
                                                                5400
 1006
        chanchal singhvi
                            director
                                        sales
                                                     03/09/38
                                                                6700
 6213
        karuna ganguly
                                                     05/06/62
                            q.m.
                                        accounts
                                                                6300
 1265
        s.n. dasqupta
                                        sales
                            manager
                                                     12/09/63
                                                                5600
 4290
       jayant choudhury
                            executive| production
                                                     07/09/50
                                                                6000
       anil aggarwal
                                       sales
 2476
                            manager
                                                     01/05/59
                                                                5000
       lalit chowdury
                            director |
                                       marketing
 6521
                                                     26/09/45
                                                                8200
        shyam saksena
 3212
                            d.g.m.
                                       accounts
                                                     12/12/55
                                                                6000
 3564
        sudhir Agarwal
                            executive| personnel
                                                     06/07/47
                                                                7500
 2345
        b. sexena
                                        marketing
                                                     12/03/45
                                                                8000
                            q.m.
        v.k.agrawal
                                        marketing
                                                                9000
                            q.m.
                                                     31/12/40
 0110 l
 adhoc@adhoc:~/Desktop$ grep -e "Agarwal" -e "aggarwal" -e "agrawal"
                                                                       emp.lst
 2476
        anil aggarwal
                            manager
                                        sales
                                                     01/05/59
                                                                5000
        sudhir Agarwal
 3564
                            executive| personnel
                                                     06/07/47
                                                                7500
        v.k.agrawal
                            g.m.
                                        marketing
 0110 I
                                                     31/12/40
                                                                9000
 adhoc@adhoc:~/Desktop$
```

-f: Taking pattern from a file

\$grep -f pattern.lst emp.lst

```
adhoc@adhoc: ~/Desktop
File Edit View Search Terminal Help
adhoc@adhoc:~/Desktop$ cat emp.lst
2233 | a.k.shukla
                                       sales
                                                    12/12/52
                                                                6000
                            q.m.
       jai sharma
                                       production
9876
                            director
                                                     12/03/50
                                                                7000
       sumit chakrobarty
                                       marketing
5678
                            d.g.m.
                                                     19/04/43
                                                                6000
       barun sengupta
                            director
                                       personnel
                                                     11/05/47
                                                                7800
2365
5423
       n.k.gupta
                            chairman I
                                       admin
                                                     30/08/56
                                                                5400
       chanchal singhvi
                                       sales
                                                     03/09/38
                                                                6700
1006
                           director |
6213
       karuna ganguly
                                       accounts
                                                     05/06/62
                                                                6300
                           g.m.
       s.n. dasgupta
1265
                            manager
                                       sales
                                                     12/09/63
                                                                5600
       jayant choudhury
                            executive| production |
                                                    07/09/50
                                                                6000
       anil aggarwal
                                       sales
2476
                            manager
                                                     01/05/59
                                                                5000
       lalit chowdury
                            director |
                                      marketing
                                                     26/09/45
                                                                8200
       shyam saksena
                            d.g.m.
                                       accounts
                                                    12/12/55
                                                                6000
3564
       sudhir Agarwal
                            executivel
                                       personnel
                                                     06/07/47
                                                                7500
2345
       j. b. sexena
                                       marketing
                                                     12/03/45
                                                                8000
                            q.m.
0110 | v.k.agrawal
                                       marketing
                                                    31/12/40
                            q.m.
                                                                9000
adhoc@adhoc:~/Desktop$ cat pattern.lst
manager
adhoc@adhoc:~/Desktop$ grep -f pattern.lst emp.lst
1265 | s.n. dasgupta
                                       sales
                                                    12/09/63
                                                                5600
                            manager
2476 | anil aggarwal
                                       sales
                                                    01/05/59
                                                                5000
                           manager
adhoc@adhoc:~/Desktop$
```

Regular Expressions

When the **pattern** have to match a specific with in the form of expression like

Then this pattern matching called Regular Expressions

Protecting Regex Metacharacters

Since many of the special characters used in regexs also have special meaning to the shell, it's a good idea to get in the habit of single quoting your regexs

- This will protect any special characters from being operated on by the shell
- If you habitually do it, you won't have to worry about when it is necessary

Symbol or Expression	Matches
*	Zero or more occurrence of previous character
g*	Nothing or g,gg,ggg,
•	A single character
*	Nothing or any number of character
[pqr]	Single character from p or q or r
[c1-c2]	Single character with range c1 to c2
[1-3]	Single digit between 1 to 3
[^pqr]	Single character which not p ,q or r
[^a-zA-Z]	Single character non alphabet

Symbol or Expression	Matches
^pat	Beginning of line
Pat\$	At end
Bash\$	Bash at end
^bash\$	Bash is only word line
^\$	Line contains nothing





| match

UNIX Tools sucks.

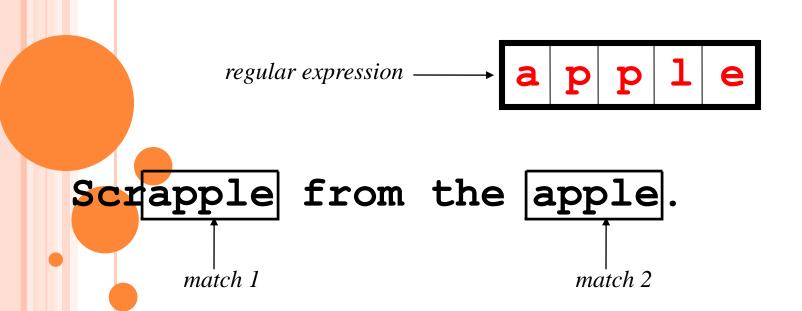
match

UNIX Tools is okay.

no match

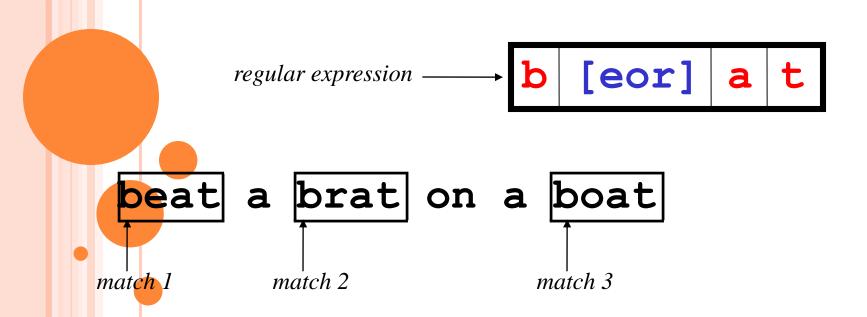
Regular Expressions

 A regular expression can match a string in more than one place.



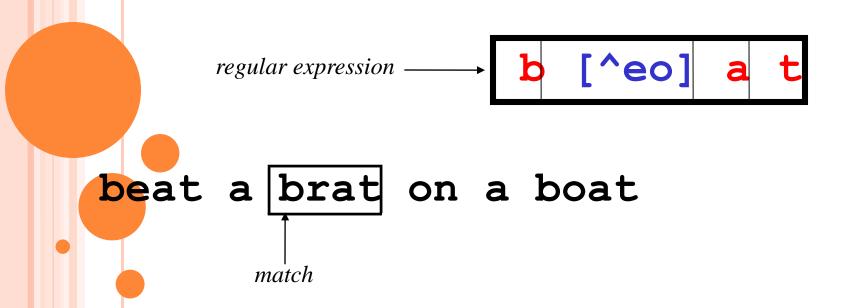
Character Classes

• Character classes [] can be used to match any specific set of characters.



Negated Character Classes

Character classes can be negated with the [^] syntax.



More About Character Classes

- [aeiou] will match any of the characters a, e, i,
 o, or u
- [kK] orn will match korn or Korn
- Ranges can also be specified in character classes
 - [1-9] is the same as [123456789]
 - [abcde] is equivalent to [a-e]
 - You can also combine multiple ranges
 - [abcde123456789] is equivalent to [a-e1-9]
 - Note that the character has a special meaning in a character class but only if it is used within a range,
 - [-123] would match the characters -, 1, 2, or 3

Named Character Classes

 Commonly used character classes can be referred to by name (alpha, lower, upper, alnum, digit, punct, cntrl)

Important for portability across languages

Anchors

• Anchors are used to match at the beginning or end of a line (or both).

• ^ means beginning of the line

\$ means end of the line

beat a brat on a boat

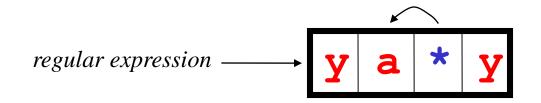
match

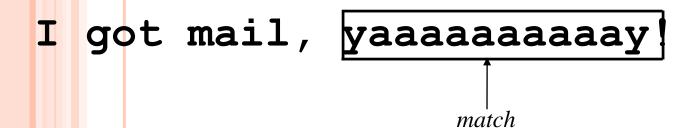
regular expression — b [eor] a t \$

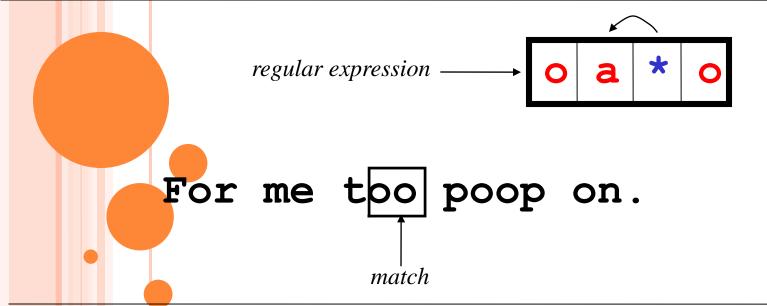
beat a brat on a boat

Repetition

• The * is used to define zero or more occurrences of the *single* regular expression preceding it.



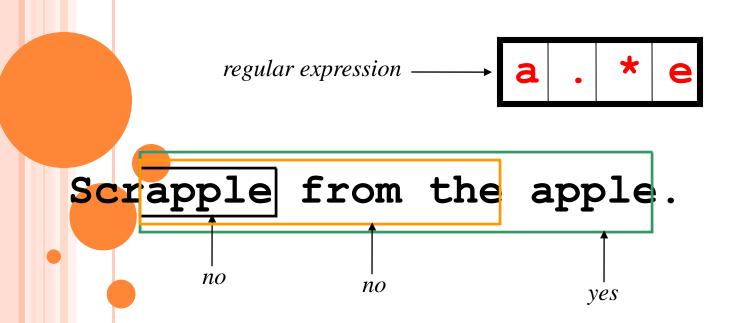






Match length

• A match will be the longest string that satisfies the regular expression.



Repetition Ranges

- Ranges can also be specified
- { } notation can specify a range of repetitions for the immediately preceding regex
- {n} means exactly n occurrences
 - n, means at least n occurrences
- { m, m} means at least n occurrences but no more than m occurrences
 - Example:
- -. {0,} same as .*
- a{2,} same as aaa*

Sub-expressions

- If you want to group part of an expression so that
 * or { } applies to more than just the previous character, use () notation
- Subexpresssions are treated like a single character
 - matches 0 or more occurrences of a
 - 🗾 aban matches ab, abc, abcc, abccc, ...
 - (abc) * matches abc, abcabc, abcabcabc, ...
 - (abc) {2,3} matches abcabe or abcabeabe

The character set

- 1. [ra]: either r or a
- 2.[aA] g [ar] [ar] wal : agaawal, agarwal, agrawal, agrrwal, Agaawal, Agarwal, Agrawal, Agrrwal

\$ grep "[aA] g [ar] [ar] wal" emp.lst

```
adhoc@adhoc: ~/Desktop
File Edit View Search Terminal Help
adhoc@adhoc:~/Desktop$ cat emp.lst
       a.k.shukla
                                      sales
2233
                                                    12/12/52
                                                               6000
                           a.m.
9876 I
      jai sharma
                           director I
                                      production
                                                    12/03/50
                                                               7000
       sumit chakrobarty |
                                      marketing
5678
                           d.g.m.
                                                    19/04/43
                                                               6000
2365
       barun sengupta
                                      personnel
                                                    11/05/47
                           director |
                                                               7800
5423
      n.k.gupta
                           chairman |
                                      admin
                                                    30/08/56
                                                               5400
       chanchal singhvi
1006
                           director |
                                      sales
                                                    03/09/38
                                                               6700
                                                    05/06/62 |
       karuna ganguly
6213
                           g.m.
                                      accounts
                                                               6300
1265
       s.n. dasgupta
                           manager
                                      sales
                                                    12/09/63
                                                               5600
       jayant choudhury
4290
                           executive| production
                                                    07/09/50
                                                               6000
       anil aggarwal
2476
                           manager
                                      sales
                                                    01/05/59
                                                               5000
      lalit chowdury
6521
                           director |
                                      marketing
                                                    26/09/45
                                                               8200
       shyam saksena
3212
                           d.q.m.
                                      accounts
                                                    12/12/55
                                                               6000
3564
      sudhir Agarwal
                           executive| personnel
                                                    06/07/47
                                                               7500
2345
      j. b. sexena
                           g.m.
                                      marketing
                                                    12/03/45
                                                               8000
0110 | v.k.agrawal
                                      marketing
                                                    31/12/40
                                                               9000
                           q.m.
adhoc@adhoc:~/Desktop$ grep "[aA]g[ar][ar]wal" emp.lst
                           executive| personnel
       sudhir Agarwal
                                                    06/07/47
3564
                                                               7500
0110 I
      v.k.agrawal
                                      marketing
                           g.m.
                                                    31/12/40
                                                               9000
adhoc@adhoc:~/Desktop$
```

The *(Immediately preeding character)

 g^* : Zero or many number of g. e.g – [aA] gg* [ar][ar]wal $\{g,gg,gg...\}$

\$grep "[aA]gg*[ar][ar]wal" emp.lst

```
adhoc@adhoc:~/Desktop$ cat emp.lst
      a.k.shukla
                                      sales
                                                   12/12/52
                                                               6000
2233
                           q.m.
9876 | jai sharma
                           director
                                      production
                                                   12/03/50
                                                               7000
      sumit chakrobarty |
                                      marketing
                           d.g.m.
                                                   19/04/43
                                                               6000
5678 I
      barun sengupta
                           director |
                                                   11/05/47
2365 l
                                      personnel
                                                               7800
5423 | n.k.gupta
                                      admin
                           chairman I
                                                   30/08/56
                                                               5400
       chanchal singhvi
                           director |
                                      sales
                                                   03/09/38
                                                               6700
1006 |
6213
       karuna ganguly
                           q.m.
                                      accounts
                                                    05/06/62
                                                               6300
      s.n. dasqupta
1265
                           manager
                                      sales
                                                   12/09/63
                                                               5600
4290 | jayant choudhury
                           executive| production
                                                   07/09/50
                                                               6000
2476 | anil aggarwal
                         manager
                                     | sales
                                                   01/05/59
                                                               5000
6521 | lalit chowdury
                           director |
                                      marketing
                                                   26/09/45
                                                               8200
       shyam saksena
3212 I
                                      accounts
                                                   12/12/55
                           d.g.m.
                                                               6000
       sudhir Agarwal
3564
                           executive| personnel
                                                   06/07/47
                                                               7500
      j. b. sexena
                                      marketing
                           g.m.
                                                   12/03/45
                                                               8000
0110 | v.k.agrawal
                                      marketing
                           g.m.
                                                   31/12/40
                                                               9000
adhoc@adhoc:~/Desktop$ grep "[aA]gg*[ar][ar]wal"
                                                  emp.lst
2476 | anil aggarwal
                           manager
                                      sales
                                                   01/05/59
                                                               5000
      sudhir Agarwal
                           executive| personnel
                                                   06/07/47
                                                               7500
0110 | v.k.agrawal
                                      marketing
                                                   31/12/40
                           g.m.
                                                               9000
adhoc@adhoc:~/Desktop$
```

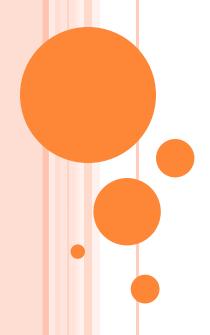
The Dot(.)

Matches a single character (like shell use? Same as here is dot) e.g

2... -> Match four character begin with 2

2??? -> Match four character begin with 2 in **SHELL**

\$grep "J.*rma" emp.lst



File Edit View Search Te	erminal Help		
adhoc@adhoc:~/Desktop!	cat emp.lst		
2233 a.k.shukla	g.m.	sales	12/12/52 6000
9876 jai sharma	director	production	12/03/50 7000
5678 sumit chakroba	rty d.g.m.	marketing	19/04/43 6000
2365 barun sengupta	director	personnel	11/05/47 7800
5423 n.k.gupta	chairman	admin	30/08/56 5400
1006 chanchal singh	/i director	sales	03/09/38 6700
6213 karuna ganguly	g.m.	accounts	05/06/62 6300
1265 s.n. dasgupta	manager	sales	12/09/63 5600
4290 jayant choudhu	ry executive	production	07/09/50 6000
2476 anil aggarwal	manager	sales	01/05/59 5000
6521 lalit chowdury	director	marketing	26/09/45 8200
3212 shyam saksena	d.g.m.	accounts	12/12/55 6000
3564 sudhir Agarwal	executive	personnel	06/07/47 7500
2345 j. b. sexena	g.m.	marketing	12/03/45 8000
0110 v.k.agrawal	g.m.	marketing	31/12/40 9000
adhoc@adhoc:~/Desktop	ş grep "j.*rma"	emp.lst	
9876 jai sharma		production	12/03/50 7000
adhoc@adhoc:~/Desktop	\$ grep "j.*" emp		
9876 jai sharma	director		
4290 jayant choudhu	ry executive	production	07/09/50 6000
2345 j. b. sexena	_ g.m.	marketing	12/03/45 8000
adhoc@adhoc:~/Desktop	\$ <u> </u>		

Specifying Pattern Locations (^ and \$)

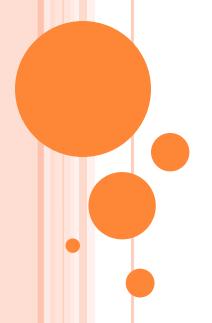
^: beginning of line

\$: end of line

e.g Show employee details whose emp_id beginning with 2

 $2 \dots$

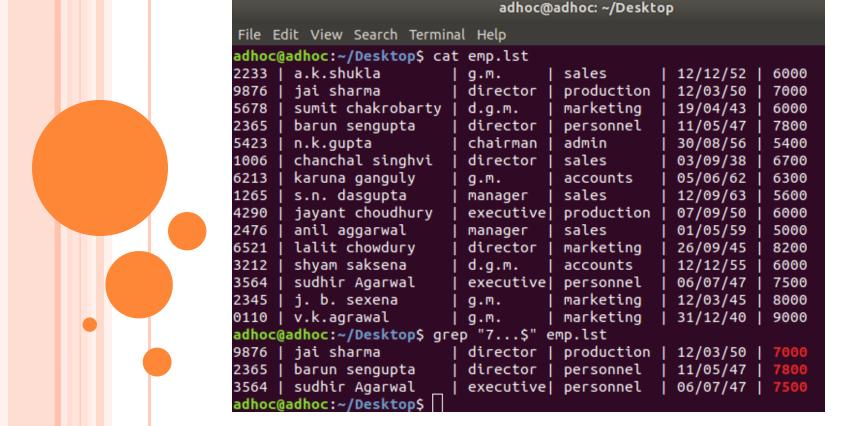
\$grep "^2" emp.lst



```
adhoc@adhoc: ~/Desktop
File Edit View Search Terminal Help
adhoc@adhoc:~/Desktop$ cat emp.lst
2233 | a.k.shukla
                                       sales
                                                   12/12/52
                                                                6000
                            q.m.
9876 | iai sharma
                            director I
                                       production | 12/03/50
                                                                7000
      sumit chakrobarty |
                                       marketing
                            d.g.m.
                                                    19/04/43
                                                                6000
      barun sengupta
                                       personnel
                                                    11/05/47
2365
                           director |
                                                                7800
       n.k.gupta
                                       admin
                            chairman l
                                                    30/08/56
                                                                5400
      chanchal singhvi
                                                    03/09/38
1006
                            director
                                       sales
                                                                6700
       karuna ganguly
6213
                            g.m.
                                       accounts
                                                    05/06/62
                                                                6300
       s.n. dasqupta
                            manager
                                       sales
                                                    12/09/63
                                                                5600
      jayant choudhury
                           executive| production |
                                                    07/09/50
                                                                6000
      anil aggarwal
2476
                                       sales
                                                     01/05/59
                                                                5000
                           manager
       lalit chowdury
                                                     26/09/45
6521
                            director
                                       marketing
                                                                8200
       shyam saksena
3212
                                       accounts
                                                    12/12/55
                                                                6000
                            d.g.m.
      sudhir Agarwal
                                                    06/07/47
                            executive| personnel
                                                                7500
      j. b. sexena
                            q.m.
                                       marketing
                                                    12/03/45
                                                                8000
0110 | v.k.agrawal
                            q.m.
                                      marketing
                                                    31/12/40
                                                                9000
adhoc@adhoc:~/Desktop$ grep
                             "^2" emp.lst
2233 | a.k.shukla
                                       sales
                            g.m.
                                                   12/12/52
                                                                6000
       barun sengupta
                                                    11/05/47
                            director
                                       personnel
                                                                7800
       anil aggarwal
                                       sales
                                                                5000
                            manager
                                                     01/05/59
2345 | j. b. sexena
                                       marketing
                                                    12/03/45
                            g.m.
                                                                8000
adhoc@adhoc:~/Desktop$
```

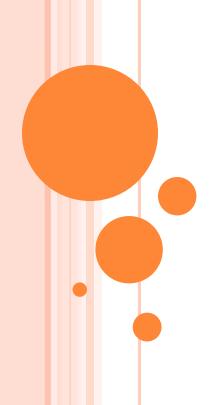
e.g Show employee details whose salary lies between 7000 and 7999

\$grep "**7...**\$" emp.lst



e.g

Show employee details whose emp_id does not start with 2 \$grep^[^2]" emp.lst



adhoc@adhoc: ~/Desktop					
File E	dit View Search Termin	al Help			
adhoc(@adhoc:~/Desktop\$ cat	t emp.lst			
2233	a.k.shukla	g.m.	sales	12/12/52	6000
9876	jai sharma	director	production	12/03/50	7000
5678	sumit chakrobarty	d.g.m.	marketing	19/04/43	6000
2365	barun sengupta	director	personnel	11/05/47	7800
5423	n.k.gupta	chairman	admin	30/08/56	5400
1006	chanchal singhvi	director	sales	03/09/38	6700
6213	karuna ganguly	g.m.	accounts	05/06/62	6300
1265	s.n. dasgupta	manager	sales	12/09/63	5600
4290	jayant choudhury	executive	production	07/09/50	6000
2476	anil aggarwal	manager	sales	01/05/59	5000
6521	lalit chowdury	director	marketing	26/09/45	8200
3212	shyam saksena	d.g.m.	accounts	12/12/55	6000
3564	sudhir Agarwal	executive	personnel	06/07/47	7500
2345	j. b. sexena	g.m.	marketing	12/03/45	8000
0110	v.k.agrawal	g.m.	marketing	31/12/40	9000
adhoc(@adhoc:~/Desktop\$ gre	ep "^[^2]" €	emp.lst		
9876	jai sharma	director	production	12/03/50	7000
5 678	sumit chakrobarty	d.g.m.	marketing	19/04/43	6000
5 423	n.k.gupta	chairman	admin	30/08/56	5400
1 006	chanchal singhvi	director	sales	03/09/38	6700
6 213	karuna ganguly	g.m.	accounts	05/06/62	6300
1 265	s.n. dasgupta	manager	sales	12/09/63	5600
4 290	jayant choudhury	executive	production	07/09/50	6000
6 521	lalit chowdury	director	marketing	26/09/45	8200
3212	shyam saksena	d.g.m.	accounts	12/12/55	6000
3 564	sudhir Agarwal	executive	P	06/07/47	7500
0110	v.k.agrawal	g.m.	marketing	31/12/40	9000
adhoc@adhoc:~/Desktop\$					

Escaping Special Characters

- Even though we are single quoting our regexs so the shell won't interpret the special characters, some characters are special to **grep** (eg * and .)
- To get literal characters, we *escape* the character with a \ (backslash)
- Suppose we want to search for the character sequence a*b*
 - Unless we do something special, this will match zero or more 'a's followed by zero or more 'b's, not what
 - we want
 - a *b * will fix this now the asterisks are treated as regular characters

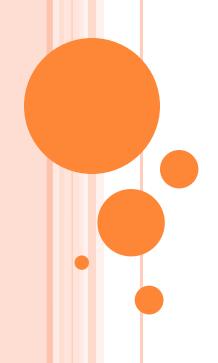
Escaping

e.g

Want to search g* in emp.lst

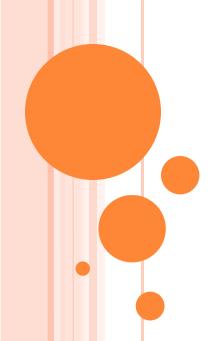
If \$ grep "g*" emp.lst :> Find all g

\$grep "\g*" emp.lst



$\mathbf{I} \mathbf{g}$					
adhoc@	adhoc:~/Desktop\$ cat	emp.lst			
2233	a.k.shukla	g.m.	sales	12/12/52	6000
9876	jai sharma	director	production	12/03/50	7000
5678	sumit chakrobarty	d.g.m.	marketing	19/04/43	6000
2365	barun sengupta	director	personnel	11/05/47	7800
5423	n.k.gupta	chairman	admin	30/08/56	5400
1006	chanchal singhvi	director	sales	03/09/38	6700
6213	karuna ganguly	g.m.	accounts	05/06/62	6300
1265	s.n. dasgupta	manager	sales	12/09/63	5600
4290	jayant choudhury	executive	production	07/09/50	6000
2476	anil aggarwal	manager	sales	01/05/59	5000
6521	lalit chowduryg*	director	marketing	26/09/45	8200
3212	shyam saksena	d.g.m.	accounts	12/12/55	6000
3564	sudhir Agarwal	executive	personnel	06/07/47	7500
2345	j. b. sexena	g.m.	marketing	12/03/45	8000
0110	v.k.agrawal	g.m.	marketing	31/12/40	9000
adhoc@	adhoc:~/Desktop\$ gre	ep "g*" emp.	lst		
2233	a.k.shukla	g.m.	sales	12/12/52	6000
9876	jai sharma	director	production	12/03/50	7000
5678	sumit chakrobarty	d.g.m.	marketing	19/04/43	6000
2365	barun sengupta	director	personnel	11/05/47	7800
5423	n.k.gupta	chairman	admin	30/08/56	5400
1006	chanchal singhvi	director	sales	03/09/38	6700
6213	karuna ganguly	g.m.	accounts	05/06/62	6300
1265	s.n. dasgupta	manager	sales	12/09/63	5600
4290	jayant choudhury	executive	production	07/09/50	6000
2476	anil aggarwal	manager	sales	01/05/59	5000
6521	lalit chowduryg*	director	marketing	26/09/45	8200
3212	shyam saksena	d.g.m.	accounts	12/12/55	6000
3564	sudhir Agarwal	executive	personnel	06/07/47	7500
2345	j. b. sexena	g.m.	marketing	12/03/45	8000
0110	v.k.agrawal	g.m.	marketing	31/12/40	9000
adhoc@adhoc:~/Desktop\$ grep "g*" emp.lst					
6521 lalit chowduryg*_ director marketing 26/09/45 8200					
adhoc@	adhoc:~/Desktop\$				

Extended R.E



Egrep: Alternation

- Regex also provides an alternation character | for matching one or another subexpression
 - (T|F1) an will match 'Tan' or 'Flan'
 - ^(From|Subject): will match the From and Subject lines of a typical email message
 - It matches a beginning of line followed by either the characters 'From' or 'Subject' followed by a ':'
- Subexpressions are used to limit the scope of the alternation
 - At (ten|nine) tion then matches "Attention" or "Atninetion", not "Atten" or "ninetion" as would happen without the parenthesis Atten|ninetion

Egrep: Repetition Shorthands

- The * (star) has already been seen to specify zero or more occurrences of the immediately preceding character
- + (plus) means "one or more"
 - abc+d will match 'abcd', 'abccd', or 'abcccccd'
 - but will not match 'abd'
 - Equivalent to {1,}

Egrep: Repetition Shorthands

- The '?' (question mark) specifies an **optional** character, the single character that immediately precedes it
 - July? will match 'Jul' or 'July'
 - Equivalent to {0,1}
 - Also equivalent to (Jul | July)
 - The *, ?, and + are known as *quantifiers* because they specify the quantity of a match
 - Quantifiers can also be used with subexpressions
 - (a*c) + will match 'c', 'ac', 'aac' or 'aacaacac' but will not match 'a' or a blank line

\$grep -E "[aA]gg?arwal" emp.lst

```
adhoc@adhoc: ~/Desktop
File Edit View Search Terminal Help
adhoc@adhoc:~/Desktop$ cat emp.lst
       a.k.shukla
                                      sales
                                                   12/12/52
2233 l
                           q.m.
                                                              6000
      jai sharma
9876
                           director
                                      production
                                                   12/03/50
                                                              7000
       sumit chakrobarty |
                                      marketing
5678
                           d.q.m.
                                                   19/04/43
                                                              6000
       barun sengupta
                           director
                                      personnel
2365
                                                   11/05/47
                                                              7800
      n.k.gupta
                           chairman I
                                      admin
5423
                                                   30/08/56
                                                              5400
       chanchal singhvi
                           director | sales
                                                   03/09/38 |
1006
                                                              6700
       karuna ganguly
6213
                         q.m.
                                    l accounts
                                                   05/06/62 |
                                                              6300
      s.n. dasqupta
                                                   12/09/63 |
1265
                         manager
                                     sales
                                                              5600
      jayant choudhury
                           executive| production
                                                   07/09/50
4290
                                                              6000
      anil aggarwal
                                    | sales
                                                   01/05/59
2476
                         manager
                                                              5000
      lalit chowduryg*
                         | director | marketing
6521
                                                   26/09/45
                                                              8200
       shyam saksena
                          d.g.m.
3212
                                      accounts
                                                  12/12/55
                                                              6000
       sudhir Agarwal
                           executive| personnel
                                                   06/07/47
3564
                                                              7500
       i. b. sexena
                                      marketing
                           g.m.
                                                   12/03/45
                                                              8000
                                      marketing
      v.k.agrawal
                                                   31/12/40
0110 l
                           q.m.
                                                              9000
adhoc@adhoc:~/Desktop$ grep -E "[aA]gg?arwal" emp.lst
       anil aggarwal
                           manager
                                                   01/05/59
2476
                                      sales
                                                              5000
      sudhir Agarwal
                           executive| personnel
                                                   06/07/47
                                                              7500
adhoc@adhoc:~/Desktop$
```

```
#inclue_<stdio.h>, #inclue_ _<stdio.h>, #inclue_ _ _<stdio.h>,
#inclue_ _ _ _ <stdio.h>
:> #incluide_+<stdio.h>
```

Matching multiple patterns(|,())

\$grep -E 'Kishan | Ravikishan" emp.lst

\$grep -E "(| Ravi)kishan" emp.lst

```
adhoc@adhoc: ~/Desktop
File Edit View Search Terminal Help
adhoc@adhoc:~/Desktop$ cat emp.lst
      a.k.shukla
                                      sales
2233 l
                                                   12/12/52
                                                               6000
                           q.m.
9876
      jai sharma
                                      production |
                           director |
                                                    12/03/50
                                                               7000
5678
       sumit chakrobarty
                           d.g.m.
                                      marketing
                                                    19/04/43
                                                               6000
       barun sengupta
                           director I
                                      personnel
                                                               7800
2365
                                                    11/05/47
      n.k.gupta
                           chairman
                                      admin
                                                               5400
5423
                                                    30/08/56
       chanchal singhvi
1006
                           director |
                                      sales
                                                    03/09/38
                                                               6700
6213
       karuna ganguly
                                      accounts
                                                    05/06/62
                                                               6300
                           q.m.
      s.n. dasqupta
                                      sales
1265
                                                    12/09/63
                                                               5600
                           manager
      jayant choudhury
                           executive| production |
4290
                                                    07/09/50
                                                               6000
2476 | anil aggarwal
                           manager
                                      sales
                                                    01/05/59
                                                               5000
      lalit chowduryg*
                           director |
                                      marketing
                                                    26/09/45
6521
                                                               8200
3212
      shyam saksena
                           d.g.m.
                                      accounts
                                                    12/12/55
                                                               6000
       sudhir Agarwal
                           executive
                                      personnel
3564
                                                    06/07/47
                                                               7500
       j. b. sexena
                                      marketing
                                                   12/03/45
                                                               8000
                           g.m.
0110 | v.k.agrawal
                                      marketing
                                                   31/12/40
                           q.m.
                                                               9000
adhoc@adhoc:~/Desktop$ grep -E "(sen|das)gupta"
                                                 emp.lst
      barun sengupta
                           director
                                      personnel
                                                    11/05/47
                                                               7800
2365 l
1265 | s.n. dasgupta
                                       sales
                                                    12/09/63
                                                               5600
                           manager
adhoc@adhoc:~/Desktop$
```

Practical Regex Examples

- Variable names in C
 - $[a-zA-Z_{]}[a-zA-Z_{0}-9]*$
- Dollar amount with optional cents
 - \\$[0-9]+(\.[0-9][0-9])?
- Time of day
 - -(1[012]|[1-9]):[0-5][0-9] (am|pm)
- HTML headers <h1> <H1> <h2> ...
 - -<[hH][1-4]>

grep Examples

```
grep 'men' File_name
grep 'fo*' File_name
egrep 'fo+' File_name
egrep -n '[Tt]he' File_name
fgrep 'The' File_name
egrep 'NC+[0-9]*A?' File_name
fgrep -f expfile File_name
```

• Find all lines with signed numbers

```
$ egrep '[-+][0-9]+\.?[0-9]*' *.c
bsearch. c: return -1;
strcmp. c: return -1;
strcmp. c: return +1;
```

• egrep has its limits: For example, it cannot match all lines that contain a number divisible by 7.

This is one line of text o.*o

input line regular expression

X	Ordinary characters match themselves	
	(NEWLINES and metacharacters excluded)	
xyz	Ordinary strings match themselves	
\m	Matches literal character <i>m</i>	
^	Start of line	
\$	End of line	
	Any single character	
$[xy^{\$}x]$	Any of x, y, ^, \$, or z	
[^xy^\$z]	Any one character other than x, y, ^, \$, or z	
[a-z]	Any single character in given range	
r*	zero or more occurrences of regex r	
r1r2	Matches r1 followed by r2	
\(r\)	Tagged regular expression, matches r	
\n	Set to what matched the <i>n</i> th tagged expression	
	(n = 1-9)	
$\{n,m\}$	Repetition	
r+	One or more occurrences of r	
r?	Zero or one occurrences of r	
r1 r2	Either r1 or r2	
(r1 r2)r3	Either r1r3 or r2r3	
(r1 r2)*	Zero or more occurrences of r1 r2, e.g., r1, r1r1,	
	r <mark>2</mark> r1, r1r1r2r1,)	
{n,m}	Repetition	

fgrep, grep, egrep

grep, egrep

grep

Quick Reference