awk/grep commands

Dr. Sanjeet Kumar Nayak

August 13, 2021

- awk
  - awk is given after the developer names Aho, Weinberger, Kernighan.
  - awk is a scripting langauge for manipulating data and to generate reports.
  - It is a utility enables programmer to write tiny but effective programs.
  - It is most commonly used for pattern scanning and processing.
  - It searches one or more files for pattern matching to perform required actions.
  - awk command programming language requires no compiling, and allows the user to use variables, numeric functions, string functions, and logical operators.

Contd . . .

- awk options
  - Scan file line by line.
  - splits each input line into fields.
  - compares input lines/fileds to patterns.
  - performs actions on matched lines.
- awk is used for
  - Transform Data files
  - Produce Formatted Reports
- awk programming constraints
  - Format Output Lines
  - Arithmetic and String Operations
  - Conditions and Loops
- Syntax

  awk options 'selection_criteria {action}' input-file > output-file

awk Example

- **employee.txt**

```
101   santosh      GITAM    CSE    35000
102   mahendra     XYZ      CSE    34200
103   prateek      MNP      CSE    36000
104   kiruthika    YYY      CSE    35550
105   mohanapriya  AAA      CSE    38000
106   Joyashree    MTH      MAT    39000
107   Snigdha      SSS      MAT    34000
108   Mercy        GGG      CSE    35000
109   Murali       Vignan   ECE    35000
110   Pavan        SSS      Mech   34000
111   Vijay        IIITDM   EC     34700
112   pradeep      SCHL     BED    65000
113   Aruna        SCHL     BED    65000
114   Ramya        MBL      MB     43000
```

- Print the row that matches Pattern "santosh"

  ```
  awk '/santosh/ print' employee.txt
  ```

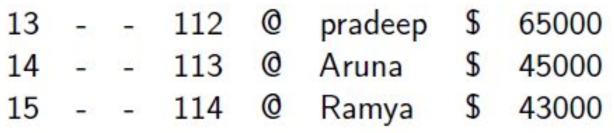
  Output

  ```
  101    santosh    GITAM    CSE    35000
  ```

- Prints all the employee details whose salary is 35000 with line numbers

  ```
  awk 'NR==13, NR==15 {print NR,"-",$1,"@",$2,"$",$NF}'
  employee.txt
  ```

  Output



```
13  -  -  112  @  pradeep  $  65000
14  -  -  113  @  Aruna    $  45000
15  -  -  114  @  Ramya    $  43000
```

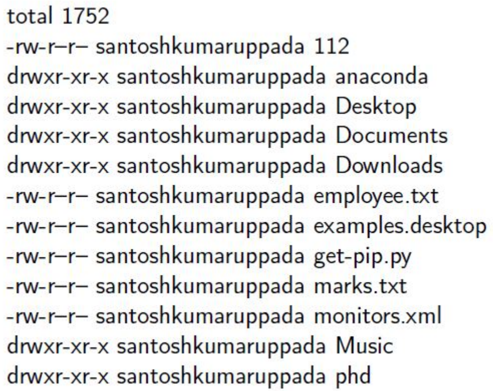
- Print length of the maximum row in the specified range
  ```
  awk 'NR==12, NR==14
  {if(length($0) > max)
  max=length($0)}
  END {print max}' employee.txt
  ```
  Output
  ```
  26
  ```
- Prints count of Rows whose length is greater than 28
  ```
  awk 'length($0)>28 {print "Row_length > 28 = " NR
  "Rows"}' employee.txt
  ```
  Output
  ```
  Row_length > 28 = 5 Rows
  ```

Contd . . .

- print First, Third and Last column of ls Long listing format

```
ls -l—awk '{print $1,$3,$NF}'
```

Output



```
total 1752
-rw-r–r– santoshkumaruppada 112
drwxr-xr-x santoshkumaruppada anaconda
drwxr-xr-x santoshkumaruppada Desktop
drwxr-xr-x santoshkumaruppada Documents
drwxr-xr-x santoshkumaruppada Downloads
-rw-r–r– santoshkumaruppada employee.txt
-rw-r–r– santoshkumaruppada examples.desktop
-rw-r–r– santoshkumaruppada get-pip.py
-rw-r–r– santoshkumaruppada marks.txt
-rw-r–r– santoshkumaruppada monitors.xml
drwxr-xr-x santoshkumaruppada Music
drwxr-xr-x santoshkumaruppada phd
```

- grep stands for globally search regular expression and print out.
- It is a command utility for searching plain-text data sets for lines that match regular expression.
- The pattern that is searched in a file is termed as regular expression.

Option	Description
-c	Count of lines that match a pattern
-h	Display matched lines, but don't display its filenames
-i	Ignore case for matching
-l	Display list of filenames only
-n	Display matched lines and line numbers
-v	Prints lines that do not match the pattern
-w	Match whole word, Ignore substring
-o	print only matched part of matching line
-A n	prints n lines after match
-B n	Prints n lines before match
-C n	Prints n lines above and below match

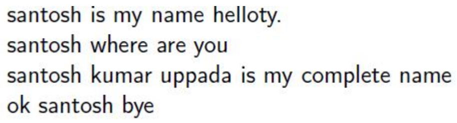
Sample data

sample.txt hello hello hello hello. hello hello hello hello.
hai hello hai hello hai. hello hai hello hai hello.
I want to add a comment here
hai hai hai hai hai hai hai hai hai hai hai hai hai hai.
hello hello hello hello. hello hello hello hello hello hello.
another comment
hai hello hai. hello he is good boy.
santosh is my name helloty.
santosh where are you.
santosh kumar uppada is my complete name.
ok santosh bye.
hi to sleep.
HAI HAI HAI HAI HAI HAI.
hai

sample1.txt Hai my name is santosh.
Hello am i doing it correctly.
hai i am going to work on grep.
we should have grip on grep.
hello did you get my point.
i said hai long back but no one wished me hai till now.
so no hai from now i will say only bye.
hello bye
santosh signing-off
SANTOSH KUMAR UPPADA
SANSANTOSH KUMKUMAR UPPUPPADA

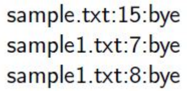
Examples

- Print only matched lines, with ignored case from specified files sample.txt, sample1.txt
  grep -hi "bye" sample.txt sample1.txt
  Output



```
santosh is my name helloty.
santosh where are you
santosh kumar uppada is my complete name
ok santosh bye
```

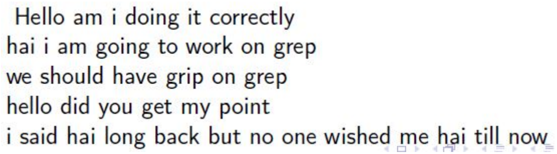
- Prints only matched words, ignoring substring with ignored case from specified files sample.txt, sample1.txt with its line numbers
  grep -woni "bye" sample.txt sample1.txt
  Output



```
sample.txt:15:bye
sample1.txt:7:bye
sample1.txt:8:bye
```

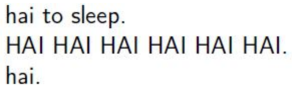
Contd . . .

- Prints the lines that do not match with the specified pattern
  grep -vi "santosh" sample1.txt
  Output



```
 Hello am i doing it correctly
hai i am going to work on grep
we should have grip on grep
hello did you get my point
i said hai long back but no one wished me hai till now
```

- Prints two lines next to the matching pattern in addition to printing matching line
  grep -A 2 "sleep" sample.txt
  Output



```
hai to sleep.
HAI HAI HAI HAI HAI HAI.
hai.
```

Conclusion

- fgrep stands for fixed-character grep.
- fgrep is generally used when dealing with files that contains more meta-characters.
- It is similar of using grep with f option. It was termed to be Faster as it does not interpret regular expression. grep -f
- rgrep stands for recursive grep which finds patterns from all sub-folders and files in the present working directory.
- rgrep is similar to using grep -r
- Other variant of grep is zgrep,which is used for finding patterns from compressed files.