COMS 6100 - Homework 8 - Linux - Regular Expressions

September 29, 2020

1. **Command**: grep --color=auto ing 161.txt

   - **Prediction:** This command will search for "ing" and color it.
   - **Section of actual output**

     The Foundation is committed to complying with the laws regulating
     against accepting unsolicited donations from donors in such states who
     any statements concerning tax treatment of donations received from
     ways including including checks, online payments and credit card
     including how to make donations to the Project Gutenberg Literary

   - **Analysis:** It looks in the file words containing "ing" and returns
     the whole line. It also highlights in "ing" in each line.

2. **Command**: grep --color=auto ˆB 161.txt

   - **Prediction:** This command will search for anything that starts
     with a capital B.
   - **Section of actual output**

     But so little interest had he taken in the matter, that he owed all his
     Brandon seems a great deal at home, nobody can tell what may
     But though Mrs. Ferrars DID come to see them, and always treated them
     But so it was. Instead of falling a sacrifice to an irresistible
     Between Barton and Delaford, there was that constant communication

   - **Analysis:** This command looks for any lines starting with a
     capital B. It then highlights the B in red.

3. **Command**: grep --color=auto l$ 161.txt

- **Prediction:** This command will look for any line that ends in l, highlight the l, and return each line.

- **Section of actual output**

  ```
  liability to you for damages, costs and expenses, including legal
  goals and ensuring that the Project Gutenberg-tm collection will
  state of Mississippi and granted tax exempt status by the Internal
  809 North 1500 West, Salt Lake City, UT 84116, (801) 596-1887, email
  information can be found at the Foundation's web site and official
  ```

- **Analysis:** This looked for any line that ends in l, highlighted the l, and returned each line.

4. **Command**: grep --color=auto "\.$" 161.txt

   - **Prediction:** This command will search for any line that ends in a period, highlight the period, and return the whole line.

   - **Section of actual output**

     ```
     works.
     Gutenberg-tm eBooks with only a loose network of volunteer support.
     editions, all of which are confirmed as Public Domain in the U.S.
     keep eBooks in compliance with any particular paper edition.
     subscribe to our email newsletter to hear about new eBooks.
     ```

   - **Analysis:** This command searched for any line that ended in a period, highlighted the period, and returned the whole line.

5. **Command**: grep -v --color=auto "the" 161.txt

   - **Prediction:** Not sure on this one. It will search for the word "the" and highlight it in red but not sure about the -v.

   - **Section of actual output**

     ```
         http://www.gutenberg.net

     This Web site includes information about Project Gutenberg-tm,
     Archive Foundation, how to help produce our new eBooks, and how to
     subscribe to our email newsletter to hear about new eBooks.
     ```

   - **Analysis:** Returns all lines that does not have a "the" in it.

6. **Command**: grep -v --color=auto " the " 161.txt

- **Prediction:** This command will search for sentences that have the string " the ". including the spaces, and return all other lines that do not match this criteria.

- **Section of actual output**

  ```
      http://www.gutenberg.net

  This Web site includes information about Project Gutenberg-tm,
  Archive Foundation, how to help produce our new eBooks, and how to
  subscribe to our email newsletter to hear about new eBooks.
  ```

- **Analysis:** This command searched for sentences that have the string " the ". including the spaces, and returned all other lines that did not match this criteria.

7. **Command**: grep --color=auto ".oo.ness" 161.txt

   - **Prediction:** This command will search for any line that has "oo" then "ness" in a line. highlight the word, then return the line.

   - **Section of actual output**

     ```
     and I feel the goodness of Colonel Brandon most sensibly.  There are
     for I cannot be ignorant that to you, to your goodness, I owe it
     she hoped no coolness, no slight, would appear in their behaviour to
     his goodness--with a conviction of his fond attachment to herself,
     between themselves, or producing coolness between their husbands.
     ```

   - **Analysis:** This command searched for any line that has "oo" then "ness" in a line. highlighted the word, then returned the line.

8. **Command**: grep --color=auto T[a-z]* 161.txt

   - **Prediction:** This command will search for T followed by some letter then anything else. It will then return the lines and highlight the results.

   - **Section of actual output**

     ```
     where we have not received written confirmation of compliance.  To
     SEND DONATIONS or determine the status of compliance for any
     donations.  To donate, please visit: http://pglaf.org/donate
     unless a copyright notice is included.  Thus, we do not necessarily
     This Web site includes information about Project Gutenberg-tm,
     ```

- **Analysis:** This command searched for T followed by some letter then anything else. It returned the lines and highlighted the results.

9. **Command**: grep -E --color=auto "bo+" 161.txt

   - **Prediction:** This command will find any words that start with "bo" with characters at the end or the end of line.

   - **Section of actual output**

   ```
   Section 3.  Information about the Project Gutenberg Literary Archive
   Section 4.  Information about Donations to the Project Gutenberg
   Section 5.  General Information About Project Gutenberg-tm electronic
   This Web site includes information about Project Gutenberg-tm,
   subscribe to our email newsletter to hear about new eBooks.
   ```

   - **Analysis:** This command searched for any lines that had "bo" and returned the whole line, highlighting the results.

10. **Command**: egrep --color=auto "bo+" 161.txt

    - **Prediction:** Same as the commend before this. Not sure what egrep vs grep would do.

    - **Section of actual output**

    ```
    Section 3.  Information about the Project Gutenberg Literary Archive
    Section 4.  Information about Donations to the Project Gutenberg
    Section 5.  General Information About Project Gutenberg-tm electronic
    This Web site includes information about Project Gutenberg-tm,
    subscribe to our email newsletter to hear about new eBooks.
    ```

    - **Analysis:** Same as the commend before this. No noticeable difference.

11. **Command**: grep --color=auto the 161.txt

    - **Prediction:** Searches for the word "the" and returns lines with that word.

    - **Section of actual output**

    ```
    methods and addresses.  Donations are accepted in a number of other
    Professor Michael S. Hart is the originator of the Project Gutenberg-tm
    editions, all of which are confirmed as Public Domain in the U.S.
    Most people start at our Web site which has the main PG search facility:
    including how to make donations to the Project Gutenberg Literary
    ```

- **Analysis:** Searches for the string "the" and returns lines with that string.

12. **Command**: grep --color=auto .oo. 161.txt

    - **Prediction:** This command will search for "oo" and turn lines with that string.
    - **Section of actual output**

      ```
      Gutenberg-tm eBooks with only a loose network of volunteer support.
      Project Gutenberg-tm eBooks are often created from several printed
      keep eBooks in compliance with any particular paper edition.
      Archive Foundation, how to help produce our new eBooks, and how to
      subscribe to our email newsletter to hear about new eBooks.
      ```

    - **Analysis:** This command searched for "oo" and turned lines with that string.

13. **Command**: grep -n --color=auto .oo. 161.txt

    - **Prediction:** This command will search for the string "oo", return the lines with that string, and do something with the formating of those lines.
    - **Section of actual output**

      ```
      13012:Gutenberg-tm eBooks with only a loose network of volunteer support.
      13015:Project Gutenberg-tm eBooks are often created from several printed
      13018:keep eBooks in compliance with any particular paper edition.
      13027:Archive Foundation, how to help produce our new eBooks, and how to
      13028:subscribe to our email newsletter to hear about new eBooks.
      ```

    - **Analysis:** This command searched for the string "oo", returned the lines with that string, and gave the line number on the left.

14. **Command**: grep -n --color=auto "t[wo]o" 161.txt

    - **Prediction:** This command will search for "to" or "two" and return the lines with their line numbers on the left.
    - **Section of actual output**

      ```
      12558:in which their husbands of course took a part, as well as the frequent
      12568:his brother too little, or bringing himself too much;--and if Edward
      12600:than herself under the event of a former attachment, whom, two years
      12601:before, she had considered too old to be married,--and who still sough
      13012:Gutenberg-tm eBooks with only a loose network of volunteer support.
      ```

- **Analysis:** This command searched for "to" or "two" and returned the lines with their line numbers on the left.

15. **Command**: grep -n --color=auto "t[wo]o. " 161.txt

    - **Prediction:** This command will search for "to " or "two " and return the lines with their line numbers on the left. Spaces included.

    - **Section of actual output**

      ```
      6234:together at Barton?  The morning that we parted too!  When he told me
      6394:finishing her wine!  And the dried cherries too!  Lord! nothing seems
      8982:too.  But why should not I wear pink ribbons?  I do not care if it IS
      11020:day or two.  Good bye."
      11924:she wished that she had shaken hands with him too.  But it was then to
      ```

    - **Analysis:** This command searched for "to" or "two " with punctuation being allowed after the string, and returned the lines with their line numbers on the left.

16. **Command**: grep -n --color=auto "t[wo]o[^k] " 161.txt

    - **Prediction:** This command will search for "to", "two", or "tok" the return results with a line number in front of the results.

    - **Section of actual output**

      ```
      11931:took a seat as far from him as she could, and maintained a strict
      11935:awful pause took place.  It was put an end to by Mrs. Dashwood, who
      11962:knowing what to do; took up a pair of scissors that lay there, and
      12467:till every thing was ready, and the ceremony took place in Barton
      12558:in which their husbands of course took a part, as well as the frequent
      ```

    - **Analysis:** The command searched for "too", "took", and "two" and returned the results with the line number on the left.

17. **Command**: grep -i --color=auto most 161.txt

    - **Prediction:** This command will search for "most" with some kind of indentation.

    - **Section of actual output**

      ```
      conduct, her most favourite maxims.  She was born to overcome an
      as the least considerable, that though sisters, and living almost
      things that you can do with most Project Gutenberg-tm electronic works
      ```

```
what you can do with this work.  Copyright laws in most countries are in
almost no restrictions whatsoever.  You may copy it, give it away or
Most people start at our Web site which has the main PG search facility:
```

- **Analysis:** This command will search for "most" and list the
  results regardless if upper or lower case.

18. **Command**: grep -i --color=auto ^most 161.txt

    - **Prediction:** This command will search for "most" regardless if
      upper or lower case.

    - **Section of actual output**

      ```
      most complete thing of its kind, he said, that ever was, and you all
      most gratifying to her feelings!  It was an advantage that could not be
      most immoderately.  The idea of Edward's being a clergyman, and living
      most wearing anxiety seemed to make requisite, was kept off by
      Most people start at our Web site which has the main PG search facility:
      ```

    - **Analysis:** This command searched for lines starting with "most".

19. **Command**: grep --color=auto "n.n." 161.txt

    - **Prediction:** This command will look for any strings that match
      n(some character)n(character).

    - **Section of actual output**

      ```
      Volunteers and financial support to provide volunteers with the
      and permanent future for Project Gutenberg-tm and future generations.
      The Project Gutenberg Literary Archive Foundation is a non profit
      ($1 to $5,000) are particularly important to maintaining tax exempt
      any statements concerning tax treatment of donations received from
      ```

    - **Analysis:** This command looked for any strings that matched
      n(some character)n(character).

20. **Command**: grep --color=auto "n.n.*" 161.txt

    - **Prediction:** This command will search for n(character)n(character)(punctuation).

    - **Section of actual output**

      ```
      Volunteers and financial support to provide volunteers with the
      and permanent future for Project Gutenberg-tm and future generations.
      The Project Gutenberg Literary Archive Foundation is a non profit
      ($1 to $5,000) are particularly important to maintaining tax exempt
      any statements concerning tax treatment of donations received from
      ```

- **Analysis:** This command will search for n(character)n(character)(punctuation) and highlight the n(character)n-and everything after that.

21. **Command**: grep --color=auto "n..n." 161.txt

   - **Prediction:** This command will look for any strings that match n(some character)(character)n(character).

   - **Section of actual output**

   ```
   Project Gutenberg-tm depends upon and cannot survive without wide
   increasing the number of public domain and licensed works that can be
   ways including including checks, online payments and credit card
   with anyone.  For thirty years, he produced and distributed Project
   editions, all of which are confirmed as Public Domain in the U.S.
   ```

   - **Analysis:** This command searched for any strings that match n(some character)(character)n(character).

22. **Command**: grep --color=auto "n[ˆ][ˆ]n." 161.txt

   - **Prediction:** This will search for n(character or punctuation)(character or punctuation)n(character) and return the results.

   - **Section of actual output**

   ```
       and discontinue all use of and all access to other copies of
   receive a refund of the money (if any) you paid for it by sending a
   trademark owner, any agent or employee of the Foundation, anyone
   ways including including checks, online payments and credit card
   with anyone.  For thirty years, he produced and distributed Project
   ```

   - **Analysis:** This command searched for n(character)(character)n.

23. **Command**: grep --color=auto"ˆ\(.*\)," 161.txt

   - **Prediction:** This command will search for anything that has two backslashes and ends with a comma.

   - **Section of actual output**

   ```
   with anyone.  For thirty years, he produced and distributed Project
   editions, all of which are confirmed as Public Domain in the U.S.
   unless a copyright notice is included.  Thus, we do not necessarily
   This Web site includes information about Project Gutenberg-tm,
   Archive Foundation, how to help produce our new eBooks, and how to
   ```

- **Analysis:** This command searches for anything that has a , in the line and highlights everything before.

24. **Command**: egrep --color=auto "ˆ\(.*\)," 161.txt

   - **Prediction:** Same command as regular grep command.
   - **Section of actual output**

   - **Analysis:** This command did not result.

25. **Command**: egrep --color=auto "ˆ(.*)," 161.txt

   - **Prediction:** This command will search for anything that has a comma and highlight everything before it.
   - **Section of actual output**

   ```
   with anyone.  For thirty years, he produced and distributed Project
   editions, all of which are confirmed as Public Domain in the U.S.
   unless a copyright notice is included.  Thus, we do not necessarily
   This Web site includes information about Project Gutenberg-tm,
   Archive Foundation, how to help produce our new eBooks, and how to
   ```

   - **Analysis:** This command searched for anything that has a comma and highlighted everything before it.

26. **Command**: egrep --color=auto ",\(.*\)$" 161.txt

   - **Prediction:** This command will find anything that has a command at the end of the line and highlight everything before it.
   - **Section of actual output**

   - **Analysis:** This command did not run for me.

27. **Command**: egrep --color=auto ",(.*)$" 161.txt

   - **Prediction:** The command will search for anything that has a comma then highlight everything after it.
   - **Section of actual output**

```
        with anyone.  For thirty years, he produced and distributed Project
        editions, all of which are confirmed as Public Domain in the U.S.
        unless a copyright notice is included.  Thus, we do not necessarily
        This Web site includes information about Project Gutenberg-tm,
        Archive Foundation, how to help produce our new eBooks, and how to
```

- **Analysis:** The command searched for anything that had a comma then highlighted everything after it, including the comma.

28. **Command**: egrep --color=auto ",\(.*\)\." 161.txt

- **Prediction:** This command will do nothing for me.
- **Section of actual output**

- **Analysis:** This command did nothing for me.

29. **Command**: egrep --color=auto "Th[^[:space:]]+" 161.txt

- **Prediction:** This command will search for any line that has a "th" then some character then a space and is not at the end of the line.
- **Section of actual output**

```
        Revenue Service.  The Foundation's EIN or federal tax identification
        The Foundation's principal office is located at 4557 Melan Dr. S.
        The Foundation is committed to complying with the laws regulating
        unless a copyright notice is included.  Thus, we do not necessarily
        This Web site includes information about Project Gutenberg-tm,
```

- **Analysis:** This command searched for any line that has a "th" then some character then a space and was not at the end of the line.

30. **Command**: egrep --color=auto "t[a-z][a-z]k" 161.txt

- **Prediction:** This command will search for anything that has a "t"(letter)(letter)"t".
- **Section of actual output**

```
        till every thing was ready, and the ceremony took place in Barton
        the rest followed in course.  Instead of talking of Edward, they came
        gradually to talk only of Robert,--a subject on which he had always
        in which their husbands of course took a part, as well as the frequent
        ever appeared in Robert's style of living or of talking to give a
```

- **Analysis:** This command searched for anything that has a "t"(letter)(letter)"t" and returned the line.

31. **Command**: egrep --color=auto -o "t[a-z][a-z]k" 161.txt

    - **Prediction:** This command will search for anything that has a "t"(letter)(letter)"t" and order the lines by the individual orders sorted.

    - **Section of actual output**

    ```
    took
    talk
    talk
    took
    talk
    ```

    - **Analysis:** This command searched for anything that has a "t"(letter)(letter)"t" and returned only the words one per line.

32. **Command**: egrep -o "t[a-z][a-z]k" 161.txt | sort | uniq -u

    - **Prediction:** Write about prediction here.

    - **Section of actual output**

    ```
    trok
    ```

    - **Analysis:** Search for unique words that have "t"(letter)(letter)"t" and only appears once.

33. **Command**: egrep -o "t[a-z][a-z]k" 161.txt | sort | uniq -d

    - **Prediction:** Search for unique words that have "t"(letter)(letter)"t" and return the words only descending.

    - **Section of actual output**

    ```
    tack
    talk
    task
    tock
    took
    trik
    ```

    - **Analysis:** Search for unique words that have "t"(letter)(letter)"t" and return the words, regardless of how many times each word has appeared.

34. **Command**: egrep -o "t[a-z][a-z]k" 161.txt | sort | uniq -d | wc

   - **Prediction:** Search for words that follow the same pattern of "t"(letter)(letter)"t" and return the words, sorted, with word counts.

   - **Section of actual output**

     ```
         6       6      30
     ```

   - **Analysis:** Calculates the word counts for words matching the pattern "t"(letter)(letter)"t".

35. **Command**: grep the *

   - **Prediction:** Search all text files for the string "the".

   - **Section of actual output**

     ```
     161.txt:methods and addresses.  Donations are accepted in a number of other
     161.txt:Professor Michael S. Hart is the originator of the Project Gutenberg
     161.txt:editions, all of which are confirmed as Public Domain in the U.S.
     161.txt:Most people start at our Web site which has the main PG search facil
     161.txt:including how to make donations to the Project Gutenberg Literary
     ```

   - **Analysis:** Search all text files for the string "the" and list the file on the left.

36. **Command**: grep -h *

   - **Prediction:** Grab the top line for each text file and list the file on the left.

   - **Section of actual output**

     ```
     grep: kitty: Is a directory
     grep: sampledir: Is a directory
     ```

   - **Analysis:** Searched for items in current directory that are directories.

37. **Command**: grep -oh the * | wc

   - **Prediction:** List files in current directory with the string "the" and how many times it appears in each file.

   - **Section of actual output**

```
grep: kitty: Is a directory
grep: sampledir: Is a directory
    7060    7060   28240
```

- **Analysis:** List files in current directory with the string "the" and how many times it appears in each file. It does not give the name of the file.

---

https://www.gutenberg.org/files/46/46-0.txt