Part#1: Regular Expression (RegEx):

* هو مجموعة من chars الي بتساعدنا نبحث عن محتوى او اي نمط سواء كان معقد او بسيط
* وبما انه بدنا نبحث عن محتوى او نمط ف رح نستخدم **grep**
* تجربة اليوم مليانه grep، للتذكير تكتب هكذا grep "pattern/word" file\_name
* رح نتعلم في هذه التجربة هاي chars ونشوف ايش معناهم وكيف منستخدمهم
* نقووول **بسم الله** ونبلششش 😉
* رح استخدم كلمة **ماتش** او **match** كثير ومعناها بكون **تطابق**

رح اشتغل على نفس محتوى الملف الموجود بالمانيوال



1- Period or Dot (.):

* بتعمل ماتش لاي حرف بغض النظر عن ايش هو الحرف ما عدا new line char

 

* عملت ماتش لكل الملف لانه بكل بساطة كل الملف عبارة عن char حتى الرقم يعتبر char عشان هيك بحكولها wildcard pattern
* وبقدر احدد مثلا انه بدي اعمل ماتش لاي شي فيه حرف a ويتبعه اي حرف

 

* وهان بدي اكون عارف اني ببحث عن نمط pattern وليس عن كلمة word
* الفرق بينهم انه النمط ممكن يكون معه كمان اشياء ثانية يعني بكون هو substring اما الكلمة بتكون هي string لحاله يعني بكون قبله وبعد فراغ space
* طب كيف بدي ابحث عن كلمة word ، يعني مثلا بدي كل الكلمات الي فيها حرف a ويتبعها اي حرف؟ لازم استخدم الشكل التالي \<a.\>



* وفي الصورة التالية رح يبحث عن اي سطر يحتوي على **كلمة** فيها الاحرف Bel ويتبعها اي حرف

 

* من المفضل استخدام double quotation ("") مع اي شيء تريد البحث عنه
* ونستطيع ان نبحث عن كلمة بطول معين، مثلا نبحث عن الاسطر الي فيها كلمة تبدأ بحرف بحرف p وتنتهي بحرف m ويكون طولها (عدد الاحرف) 7

 

2- escape char (\):

* اذا بدك تبحث عن حرف (.)، ما بزبط ابحث عنه مباشرة زي هيك:

grep “.” exp3.txt

* لانه (.) لها معنى واستخدام خاص في RegEx، ف اذا بدنا نبحث عن (.) لازم استخدم \ والي اسمها escape char

 

3- Caret (^):

* بتعمل ماتش لبداية السطر، بستخدمها اذا مثلا بدي الاسطر الي بتبلش بشي معين.

 

* فمثلا هان عملنا ماتش لكل الاسطر الي بتبلش بحرف m

Q) print the lines that start with Dot (.)

4- Dollar Sign ($):

* تستخدم لعمل ماتش لاي شي بنهاية السطر
* مثلا الاسطر الي تنتهي ب حرف (y)

 

Q)

* print the lines that end with “man” word.
* Print the blank lines (lines contains no characters)

5- Choice of characters ( [...] ):

* تستخدم لعمل ماتش لاحد ال chars ضمن range معين
* مثلا بدي اعمل ماتش ل T او t ، ف لازم احطهم هيك [Tt]
* في المثال الاتي رح ابحث عن **كلمة** the سواء كانت هيك the او هيك The

 

* في المثال الاتي رح اعمل ماتش للاسطر الي فيهم رقم (بغض النظر عن قيمة الرقم)، [1234567890] او بشكل مختصر اكثر [0-9]

 

Q) print the lines that start with any capital letter.

* رينج الاحرف [A-Z] او [a-z] او بتقدر تدمجهم مع بعض [A-Za-z]
* [-0-9]: match a dash (-) or a digit
* []0-9]: match a right bracket or a digit
* ملاحظة مهمة: اذا استخدمنا (^) داخل range ([ ])، يعني زي هيك [^] بصير معناها انه اي شي ما عدا الموجود في range
* مثلا بدي اعمل ماتش لاي chars ما عدا الارقام => [^0-9]

 

Q) remove the line that contains digits.

6- Asterisk (\*):

* تعمل ماتش لحرف مكرر صفر او اكثر من مرة 🤔
* بستخدمها اذا في حرف معين او اي حرف متكرر مرة او اكثر
* مثلا بدي اعمل ماتش لحرف H ولازم يكون موجود عالاقل مرة واحدة :

 

* طب خلينا نشوف مثال ثاني ع محتوى ثاني مثلا بدي ابحث عن حرف a ويتبعه حرف p مرة او مرتين او 3 مرات او ……. او **صفر** مرة

=> يعني رح تستخرج كل هاي الانماط : a… او ap…. او app…. او الخ

 

* ملاحظة مهمة : اذا استخدمنا \* مع [ ] بتلغي معناها في regex:

 

* هاي بصير معناها ابحث عن الاسطر الي فيها t او \*

Q) why are all contents printed in the following image?

 

7- Matching a Precise Number of Characters: \{start,end\}:

* تستخدم لعمل ماتش لمجموعة من ال chars المتكررة ضمن range انا بحدد بدايته ونهايته او اذا بدي عدد معين بالزبط
* مثلا بدي ابحث عن الاسطر الي فيها من 1 - 8 من حرف (l)

 

* وبتقدر تحدد عدد معين مثلا بدي ابحث عن الاسطر الي فيها حرفين H

 

* وممكن احدد بداية range وما احدد النهاية يعني ما بتكون النهاية رقم محدد :

 

* ف هان حرف a لازم يكون موجود ع الاقل مرة واحدة ، بتشبه هاي "\*aa"

Q) print the lines that contains word with length = 4 .

9-Saving Matched Characters: \(...\):

* لفرضا انا بدي اعمل ماتش للاسطر الي بتبلش وبتنتهي بحرف معين او كلمة معينة
* ما بزبط اعلمها باستخدام الي تعلمناه قبل بس، فبحتاج شي يخزن الماتش وارجع اقارن اي ماتش ثاني بهذا الماتش الي خزنته 🙁
* اطلع مثلا بدي اطلع الاسطر الي بتبلش بحرف وبتنتهي بنفس الحرف ، ف لازم اجيب اول حرف واخزنه واقول بدي السطر الي بنتهي بهذا الحرف الي خزنته
* ف عشان اعمل ماتش لحرف او لاي شي واخزنه في اشي اسمه ريجستر registers
* عدد هذول الريجسترز هو 9
* عشان اقدر اخزن الماتش بستخدم هذا **(\**  pattern/word/char  **)\**
* اول ماتش بحفظه ب ريجستر 1 واذا عملت كمان ماتش بحفظه في ريجستر 2 وهكذا ….
* عشان اجيب قيمة الريجستر بحط هيك n\، وتكون n تمثل رقم الريجستر
* مثلا بدي اجيب قيمة الماتش في ريجستر 1 بكتب 1\
* طب خلينا نوخذ مثال انه نطبع الاسطر الي فيها اول حرفين متشابهين :

 

* هاي معناها جيب اول char (^) وخزنه بريجستر (\ )\ طبعا ريجستر 1 لانه اول مرة بخزن في ريجستر ولازم يكون الحرف الثاني نفس الحرف الاول الي خزنته بالريجيستر الاول 1\ .
* كمان مثال بدي اطبع السطر الي ببلش وبنتهي بنفس الحرف:

 

* معناها جيب اول حرف وخزنه بالريجستر الاول وبغض النظر عن الاحرف الي بعده .\* ولازم يكون السطر بنتهي بهذا الحرف $ 1\
* طبعا اذا استخدمت (\)\ مرتين معناها انا بستخدم بالريجستر الاول والثاني وهكذا…

Part#2: Sed command:

* بستخدمها عشان اعدل على محتوى الملف او طباعة اسطر معينة
* لا تعدل على الملف نفسه وانما تطبع على التيرمنال فقط
* لها 3 اشكال كالتالي:

 1- sed “s/old\_dat/new\_data” file\_name

 

* طبعا بتمسك الملف سطر سطر وبتعمل التعديلا
* لكن بتعمل تعديل لاول ماتش في السطر، ف مثلا اذا كانت كمان كلمة Unix في السطر الاول لا يحدث عليها التعديل الا اذا استخدمنا (global option ) g

 

2- sed -n “line numbers or word p”:

* تستخدم لطباعة الاسطر التي تحددها بالرقم او الاسطر التي تحتوي على كلمة معينة:

 

* هان طبع من سطر 1 ل سطر 5

 

* وهان طبع الاسطر التي تحتوي على Unix ك pattern
* اذا بدي word الصح استخدم regex

3- sed “line number or word d” :

* لحذف الاسطر من خلال رقمها او اذا كانت تحتوي على كلمة معينة:

 

* هنا حذف الاسطر من 1 - 3
* لحذف الاسطر التي تحتوي على كلمة معينة :

 

* وهنا حذف الاسطر التي تحتوي على Unix ك pattern
* اذا بدي اياها ك word الصح انه استخدم regex يعني هيك </word>/

sed vs tr:

* tr: تستخدم مع الاحرف
* مثلا اذا حطيت اشي زي هيك

 tr “dog” “cat”

* معناها بدل حرف d ب c وحرف o ب a وحرف g ب t ، وليس ال pattern كاملة
* اذا كان قصدك تبدل pattern الي هي dog ب cat لازم تستخدم sed
* وانتبه انه حكينا pattern مش word ، اذا بدك word لازم تستخدم regex

**Task:**

**Print the lines that contains the following word pattern:**

**Word1 ……. Word2 …. Word1 …. Word2**

**Note: create your file content**

**Good Luck 😀**