CMPE346- Natural Language Processing
Classwork01

Regular Expression	Sentences in the regular language
abc	abc
a[bc]d	abd acd
a[b-k]l	abl acl .... akl
ab*c	ac abc abbc abbbc ...
a[bc]*d	ad abd acd abbd abbbd accd abcbcbbccd ...
a\[c	a[c
a[^a-z]z	a0z a1z a.z a@z "a z" ...
a.z	abz "a z" acc a0c ...
a.*z	az aaz aaaaaz abz abcdz a01bcdefz ...
a\.\]\\	a.]\
a[[:alnum:]]z	aaz abz ... a0z ... aAz aBz ...
^abc	"abc" at the beginning of the line
^abc$	a line that contains only "abc"
^[0-9]*$	a line that contains only digits
ab?c	ac abc
ab+c	abc abbc abbbc ...
ab{2}c	abbc
ab{2,}c	abbc abbbc abbbbc ...
ab{2,3}c	abbc abbbc
(abc)|(def)	abc def

Assume that we have a file a.txt and it includes the following list of words: ['she', 'sells', 'sea', 'Shells', 'by', 'the', 'sea', 'Shore']. Write a Regular expression to perform the following tasks:

a. Print all the words beginning with **sh or Sh**

b. Print all the words **longer than four** characters

c. Print all the words not beginning with **s**

d. Print all the words not includes **s** character

Text Processing Commands:

Try the following commands:

cat
sort
wc
egrep
cut
uniq