CS 6320.501 - Natural Language Processing – Assignment 1

Name: Shankaranarayanan Kallidaikuruchi Ramakrishnan
NetID: sxk190109

B. Problems

1. Regular Expressions

1. the set of all alphabetic strings
   `[a-zA-Z]+`

2. the set of all alphabetic words
   `\b[a-zA-z]+\b`

3. the set of all lower case alphabetic strings ending in a b
   `[a-z]*b`

4. the set of all lower case alphabetic words ending in a b
   `\b[a-z]*b\b`

5. the set of all strings from the alphabet {“a”, “b”} such that each “a” is immediately preceded by and immediately followed by at least one “b”
   `b+(ab+)?`

6. the set of all words from the alphabet {“a”, “b”} such that each “a” is immediately preceded by and immediately followed by at least one “b”
   `\bb+a?b+\b`

7. the set of all strings from the alphabet {“a”, “b”} that form the pattern an b m where (n+m) is even; n>=0, m>=0, and (n+m)>0
   `a(aa)*b(bb)*|(aa)*(bb)+|(aa)+(bb)*`

2. Write a single regular expression for identifying social security numbers in text.

```
\b((\d{2}[1-9]|\d[1-9]\d|[1-9]\d{2})-\d{2}-([1-9]\d{3}|\d[1-
9]\d{2}|\d{2}[1-9]\d|\d{3}[1-9])|(\d{2}[1-9]|\d[1-9]\d|[1-
9]\d{2})\d{2}([1-9]\d{3}|\d[1-9]\d{2}|\d{2}[1-9]\d|\d{3}[1-9]))\b
```

3. Telephone Numbers

```
\B\+(\(([1-9]\d|\d[1-9])\)-\((\d{2}[1-9]|\d[1-9]\d|[1-9]\d{2})\)-
\((\d{2}[1-9]|\d[1-9]\d|[1-9]\d{2})\)-\(\d{4}\)|([1-9]\d|\d[1-9])-
(\d{2}[1-9]|\d[1-9]\d|[1-9]\d{2})-(\d{2}[1-9]|\d[1-9]\d|[1-9]\d{2})-
\d{4})\b
```