Zachary Shorts

Assignment 2

LING360

Trial One

My first regular expression looked like this: \b(\w+ing|\w+[oe]r|\w+ee|\w+[ae]nce|\w\*[st]ion)\b

This found 36 matches and there were 20 nominalized verbs. This would mean that the precision of the first regex is 20/36 as there were 36 matches and only 20 were accurate. This also means that the recall is 100%. This means that all of the possible nominalized verbs were found.

['Farmer', 'falling', 'her', 'her', 'getting', 'for', 'her', 'never', 'having', 'Farmer', 'getting', 'Everything', 'buying', 'Farmer', 'fence', 'identification', 'buying', 'shooting', 'her', 'after', 'shouting', 'throwing', 'leaving', 'other', 'for', 'nothing', 'asking', 'for', 'putting', 'tiger', 'tiger', 'never', 'Employee', 'investigation', 'collision', 'acceptance']

Trial Two

My second regular expression looked like this: ((?!everything|nothing)\b\w+ing\b|(?!her|for|tiger|other|never|after)\b\w+[oe]r\b|\b\w+ee\b|(?!fence)\b\w+[ae]nce\b|\b\w\*[st]ion\b)

This utilized the negative look ahead feature of regex to take out some of the common matches that were not supposed to be there. I used this to get rid of her, for, never, everything, fence, after, other, nothing, and tiger.

This left me with a precision of 20/20 and a recall of 20/20.

['Farmer', 'falling', 'getting', 'having', 'Farmer', 'getting', 'buying', 'Farmer', 'identification', 'buying', 'shooting', 'shouting', 'throwing', 'leaving', 'asking', 'putting', 'Employee', 'investigation', 'collision', 'acceptance']