Title similarity Algorithm

**Input:** ct = list of titles to be checked for similarity, st = the seed title to be compared with

**Output:** similarity score for each candidate title when compared with the seed title

Initialization:

for every title tj in ct :

for every word wji in t:

posji <- do parts of speech tagging. (generate pos tags)

//create a synonym set (synset) for each word using the word and its POS tag

synsetji <- generate\_synset(wji , posji)

synsetki <-create synset for the seed title in similar way

for every title ti in ct:

scorei = sum(similarity between synsetji and synsetki)

return [list of scores for each title in ct with st]

To compute similarity between each synset(A,B):

score=0, words=0

for each ai in A:

//Get the similarity value of the most similar word in the other sentence

Best\_score=0

For each bi in B:

If(path\_similarity(ai,bi) > Best\_score):

Best\_score = path\_similairty(ai,bi)

Score = score + best\_score

Words = words +1

Return score/words