1. In class Question

Input:

Doc #1 : the dog saw John in the park.

Doc #2 : The little bear saw the fine fat trout in the rocky brook.

Doc #3 : The dog started chasing John.

Doc #4 : the little bear caught a fish in the rocky brook.

Tasks:

a) Top TF-IDF words for the above input.

(park, 0.916290)

(fine, 0.916290)

(started, 0.916290)

(John, 0.916290)

(fat, 0.916290)

(caught, 0.916290)

(chasing, 0.916290)

(fish, 0.916290)

(a, 0.916290)

(trout, 0.916290)

(John, 0.916290)

(rocky, 0.51082)

(little, 0.51082)

(saw, 0.51082)

(brook, 0.51082)

(bear, 0.51082)

(dog, 0.51082)

(in, 0.22314)

(the, 0.22314)

(The, 0.0)

b) Top TF-IDF words for the lemmatized input.

lemmatized input:

The dog see John in the park.
The little bear see the fine fat trout in the rocky brook.
The dog start chase John.
The little bear catch a fish in the rocky brook.

Top TF-IDF words for the lemmatized input:

(fine, 0.9162907)   (bear, 0.510825)

(park, 0.9162907)   (dog, 0.510825)

(fat, 0.9162907)   (in, 0.223143)

(catch, 0.9162907)   (the, 0.223143)

(fish, 0.9162907)   (The, 0.0)

(chase, 0.9162907)   (., 0.0)

(start, 0.9162907)

(a, 0.9162907)

(trout, 0.9162907)

(see, 0.510825)

(rocky, 0.510825)

(little, 0.510825)

(brook, 0.510825)

(John, 0.510825)

c) Top TF-IDF words for the n-gram based input.

NGram based input :

The dog see
dog see John
see John in
John in the
in the park.
the park. The
park. The little
The little bear
little bear see
bear see the
see the fine
the fine fat
fine fat trout
fat trout in
trout in the
in the rocky
the rocky brook
rocky brook. The
brook. The dog
The dog start
dog start chase
start chase John.

Chase John. the
John. The little
The little bear
little bear catch
bear catch a
catch a fish
a fish in
fish in the
in the rocky
the rocky brook
rocky brook.

Top TF-IDF words for the NGram based input

(. , 2.833213)

(brook, 2.42774)

(park, 2.140066)

(fish , 2.140066)

(trout, 2.140066)

(start, 2.140066)

(John , 2.140066)

(fat , 2.140066)

(John. , 2.140066)

(chase , 2.140066)

(fine , 2.140066)

(brook. , 2.140066)

(catch , 2.140066)

(a, 2.140066)

(dog, 1.734601)

(bear , 1.58045)

(see , 1.58045)

(little, 1.58045)

(rocky , 1.58045)

(in , 1.223775)