

CSCE470 Written Assignment 1 Rubric:

1. (2 points in total, 1/3 point each)

a)

freq inversely proportional to rank
words that are too frequent do not bear much information about the content

b)

significantly reducing storage requirements
significantly reducing computation time

c)

query phrases can lose meaning (example: to be, or not to be)

d)

deriving root of the words
done by chopping off prefix/suffix

e)

because it maps terms to documents, rather than the opposite

f)

phrase inquiries

2. (3 points in total, 1 point each)

In all cases, the indices can have a -1 shift. They can have started indexing from 0, but these answers are based on starting from one

1)

you: 1: <2, 4>; 2: <1, 7, 13, 19, 21, 27, 30, 33>
head: 2: <6>
feet: 2: <9>

2)

you are you: 1: <2>
head feet: none
you /2 you: 1: <2>, 2: <19>

3)

doc_id: <sent_id: <pos, ...>>
you: 1: <1: <2, 4>>, 2: <1: <1>, 2: <7>, 3: <13, 19>, 4: <21>, 5: <27, 30>, 6: <33>>

They could have incorporated the sentences like:

doc_id: <<pos, sent_id >...> (1 point)