Modify calculation of weighting function WeightTfIDf()

I know how to weight a document term matrix by term-frequency, inverse document frequency (tf-idf) using the following as my function. For the purposes of my investigation I’m looking at ngrams for n={1,2,3,4}. In this 538 post, the author normalized tf-idf scores by length of ngram, where each tf-idf score was multiplied by the length of each n gram. In python this was coded as:

# Sets the default number of threads to use

options(mc.cores=1)

MonoQuadTokenizer <- function(x) {

NGramTokenizer(x, Weka\_control(min = 1, max = 4))

}

tfidf <- function(x) {

weightTfIdf(x, normalize = FALSE)

}

a.dtm <- DocumentTermMatrix(corpus, control=list(

tokenize = MonoQuadTokenizer,

weighting = tfidf))

Is there anyway to modify the WeightTfIdf function to do what I want? Or is the best approach to compute a document term matrix then “normalize” the scores of rows that contain monograms, bigrams, trigrams, and quadragrams?

Any help is appreciated!