**A Computational Approach to Keyword Analysis**

Camila Lívio, Ph.D.

Q1: How can we compare the two data sets?

Q2: What words do you predict will be frequent in the CNN news data set?

Q3: Let's focus on the words preceding the keyword \*film\*. What can we learn from the output?

Q4: Did you learn anything new from inspecting the context in which these words appear more closely?

Q5: What words do you think will be more frequent in the BBC data set? After observing the graph, how does this figure compare to the CNN figure?

Q6: What did you learn about how the word \*film\* is used in the BBC news data set?

#####

Homework

#####

1. What have you learned about **how these keywords are used** given their context? (5 lines)
2. What are the advantages and limitations of using a **text mining/computational approach** to text analysis? (5 lines)
3. With your neighbor, briefly discuss and then write 3-4 lines about the shortcomings of the **data sets** we analyzed in class.