Mini Assignment 1: Word Clouds

Gerhard Laimer

November 15, 2025

List of Figures

1 No stemming, unigrams and max.words = 200. 2
2 Stemming, no filter for false. 3
3 Uni- and bigrams, colors, rotation. 4
4 Tri- and quadgrams, palette, remove spares. 5

1 Task

Produce four different word clouds by varying data sets and/or parameters in the R commands. Provide a one-paragraph discussion of the results for each of the four word clouds.

2 Word Clouds produced with *statements.csv*

2.1 No stemming, unigrams and max.words = 200

Figure 1: No stemming, unigrams and max.words = 200.

Word Cloud (1) was produced with setting *stemming* to FALSE and the *ngrams* argument to 1:1 to produce unstemmed unigrams. The *wordcloud* argument *max.word* was changed from 400 to 200. The idea behind this word cloud is to show a rawer version of the data without stemming and only single words. The most frequent words are, euro, growth, inflrion and area.

2.2 Stemming, no filter for false

Figure 2: Stemming, no filter for false.

Word Cloud (2) was produced with *stemming* is TURE and the *ngrams* argument 1:1 to produce temmed unigrams. The *wordcloud* argument *max.word* was kept equal to 200, like in word cloud (1). But now $cloud\_frame <- cloud\_frame[cloud\_frame\$word != "fals"$ was excluded to not filter for the stem fals. This could in general lead to troubles and the over-representation of *fasl* due to the creation process of the corpus, since some the default of settings are often equal to FALSE. This word cloud shows no over-representation of the stemm *fals*.

3 Word Clouds produced with *QandAs.csv*

3.1 Uni- and bigrams, colors, rotation

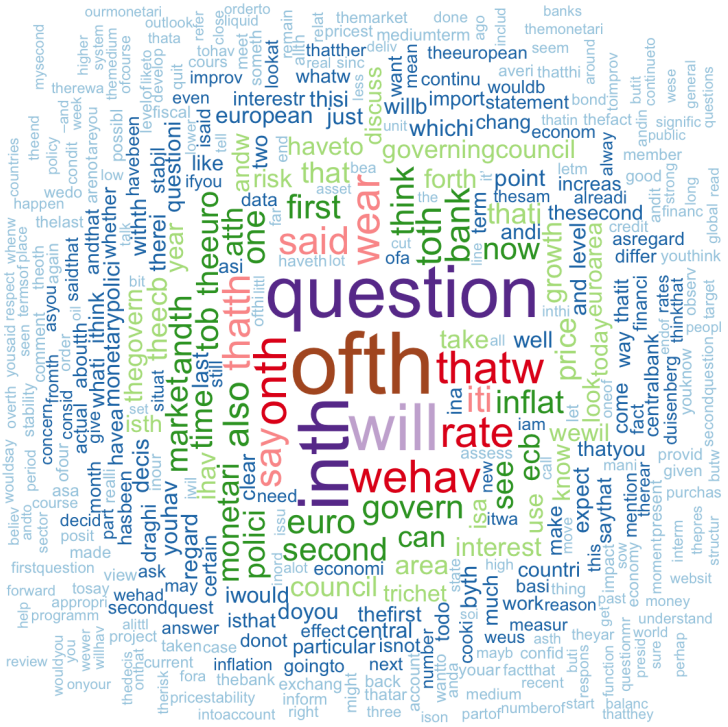

Figure 3: Uni- and bigrams, colors, rotation.

Word Cloud (3) is representative for the different visual representations of word clouds. The *max.words* were set to 400 the color palettes *colors = brewer.pal(12, "Paired")* was used, which uses up to 12 alternating complementary colors. For this plot uni- and bigrams are presented equaly distributed in vertical and horizontal elements with *rot.per = 0.5*.

3.2 Tri- and quadgrams, palette, remove spares

Figure 4: Tri- and quadgrams, palette, remove spares.

Word Cloud (4) only produces tri- and quadgrams to get a better understanding of the context of stem word combinations. To display these stems following palette was used:*colors = brewer.pal(9, "Set1")*. This word cloud also excludes its n-grams more aggressively compared to all three word clouds before, with setting *removeSparseTerms(dtm, 0.90)*, keeping n-grams only appearing at least in 10% of the documents.