

Lab 8: More complex text mining

Improving your assignment

Isabel Casas
icasas@deusto.es

Learning objectives

1. Learning to use tokens that represent pairs of words (bigrams) in sentiment analysis
2. Using it in your assignment data

Instructions

1. Lab 8 is a group activity, with the same groups as the assignment. However, if you arrive late to class, you will need to form a new group outside the existing ones. In this case, you must also choose a dataset different from your assignment if you wish to present and potentially earn an extra grade.
2. Each group has 45 minutes to work in Lab 8 and 5 minutes to present results (details below in **Output**).
3. The presentation order will be decided randomly in class, along with the presenter of each group.
4. Only group members present in Lab 8 may receive 5 extra points to add to their assignment grade for good work.

Output

- Your presentation (slides) should be in PDF or HTML format created with Rmarkdown. After 45 minutes of work, each group must email it to me. It should not have R code.
- The presentation should last only 5 minutes and only need 4 or 5 slides. First, you remind us of your problem shortly and then show:
 1. Count of most frequent words or/and wordcloud
 2. Count of most frequent bigrams
 3. Graph of the relationships between words
 4. Discuss whether conclusions from one-word and two-words analysis are different and, if so, how.