

Quantitative Text Analysis: Classwork 6

Dictionary Construction

Kenneth Benoit

In today's lab we will conduct a dictionary-based analysis of the movie review and amicus brief texts. For the movie reviews, we will use the wordcat sentiment dictionary available through QDA Miner. For the amicus briefs, we will construct a custom dictionary, using the search and keyword in context (KWIC) tools as aids to dictionary construction. The purpose of this lab is to allow you to explore two new sets of documents using the tools that you learned last week. Examine the texts first in an exploratory fashion — read a few completely, and use the retrieval tool to get a feel for important terms. Consider what level of stemming or use of stopwords is appropriate.

Instructions

1. Open QDA Miner, and create a new project from the movie review texts available on your S drive. Add the positive reviews first, and create a new variable for SENTIMENT, with values POS and NEG.
2. Assign the positive reviews that you have added as POS — you can speed up the variable assignment process by using the spreadsheet editor under the variables menu. Hold the CTRL key to select multiple cells and paste in the value you need. You can then add the negative reviews and assign the NEG value to them in the same way.
3. Open Wordstat, and select the Dictionary tab. Select the Wordcat sentiment dictionary.
4. Explore the distribution of words among dictionary categories in the Frequency tab. Experiment by altering the level of hierarchy used (an option on the dictionary panel). You can also examine the KWIC tool in Wordstat. How well does the dictionary handle negation? Can you find cases where it fails?
5. Create a new project with the Amicus briefs. Again, use the spreadsheet editor to quickly assign variable to the files — filenames beginning with sAR represent the Respondent, those with sAP the petitioner.
6. Examine the budget speeches using KWIC and cross-tabulation by the pro/anti budget variable. Which words are most likely to indicate a Petitioner or Respondent position? Discover dictionary words first by reading the text, and then confirm or alter your selections using the statistics available in the crosstab panel — use the chi-square statistic to examine terms that differ significantly across variables.
7. Create your own custom dictionary to group words as Petitioner or Respondent.
 - (a) Select 'new' from beside the Categorization dictionary option in WordStat.
 - (b) Save the new dictionary file in a location where you have write permissions.
 - (c) Select the file from the dropdown and choose 'add' categories to add the desired categories (petitioner and respondent).
 - (d) Select 'add' words, and enter the words that you want to group under each category.
8. For more detailed instructions on creating dictionaries and lemmatization rules, choose the help option in WordStat and see the section under 'Dictionaries page'.