

Homework 5

Collect some sample data (at least 20) from the following resources, which include the keyword “Fitness” and “Wearables”. Or a keyword you would like to choose.

Resources: Google-news, Twitter or Weibo (you can do Weibo in Chinese).

It is not recommended, but you can collect them manually from Twitter. However, if you collect them manually no grade will be reduced.

Calculate the sentiment of each record (or token), from the data source you have chosen.

Identify the theme for twenty of them, and report them in a excel sheet file. Your result excel file should include three columns:

- Tweet text or news entry.
- Sentiment scores from different libraries (Afinn, NRC, Bing).
- Theme, which is a keyword you have extracted from them. This means you should perform theme analysis manually as well.

You need to prepare a report on your tasks and findings along with a video file describing what you have done. You can copy paste your codes, its results and your description into a Word document, Python Notebook or you can use R notebook.

Your deadline for delivering this home work is written on the blackboard online. Please feel free to ask your question and prepare it for presentation for the next session.