Reflection :

I analyzed the subjectivity of articles on Wikipedia. Since Wikipedia serves as a widely used encyclopedia for the public, I expected the subjectivity to be very low. I also extract the articles with high subjectivity to see what topics on Wikipedia would have high subjectivity.

First, I downloaded the text from a bunch of Wikipedia articles into a file, and dump the titles to another file. Next, I used function sentiment to get the subjectivity of all the articles, and calculate the average subjectivity which turned out to be 0.37. This is a relatively low subjectivity as I expected. Then, I choose 0.5 as a standard line to determine whether a article is highly subjective or not. Finally, I traced all the highly subjective article back to their titles.

When I was doing the traceback part, I thought of implementing a function that can use the content of article to trace back to its title, because I only dumped the contents into my file. Then I realized that it would be much easier to dump all the titles in another file and use a for loop to trace the index of highly subjective articles.

Here are the titles of highly subjective articles (the values of subjectivity are in the parentheses: 'Awesome'(0.53), 'Good Day, Fellow' 'Axe Handle!'(0.51), 'I Luv You HB' Demos(0.63), 'Mad' Mike Whiddett(0.55), 'Ode-toNapoleon' hexachord(0.50).

I looked them up on Wikipedia, and find out that none of them is a formal expression, and all of these articles are relatively shorter than the other articles in general, which makes them more possible to have a high subjectivity. I concluded that Wikipedia is doing well on keeping its articles objective.

In this project, one thing I would like to improve is to make the scope larger. I wish I could include more articles from wikipedia. And also it would be better if I can eliminate some weird expressions in the articles to minimize their effect on subjectivity