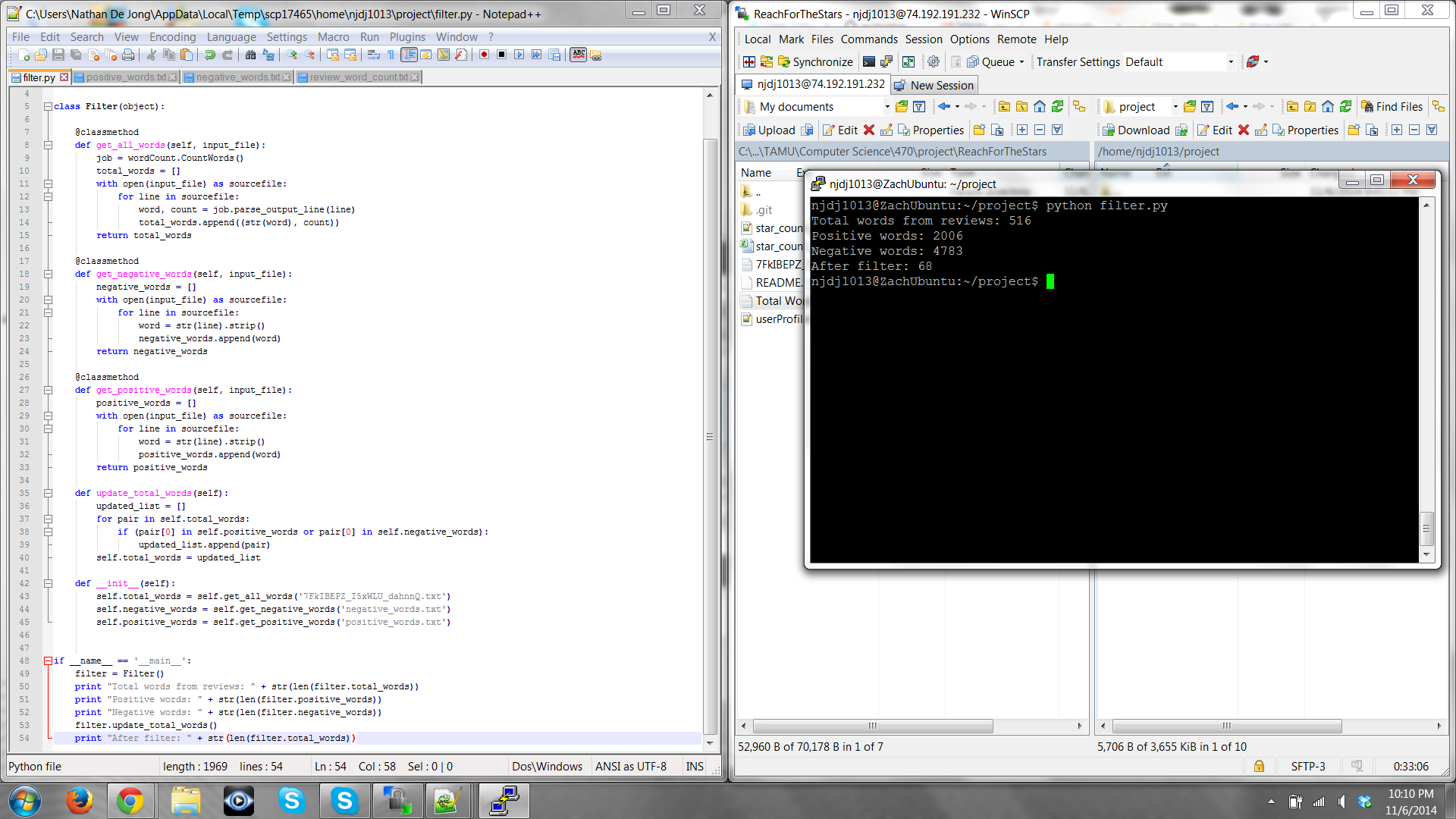
**Data Checkpoint**

For our data checkpoint we provide a few different manipulations of the data that will ultimately help us in our final classifier.

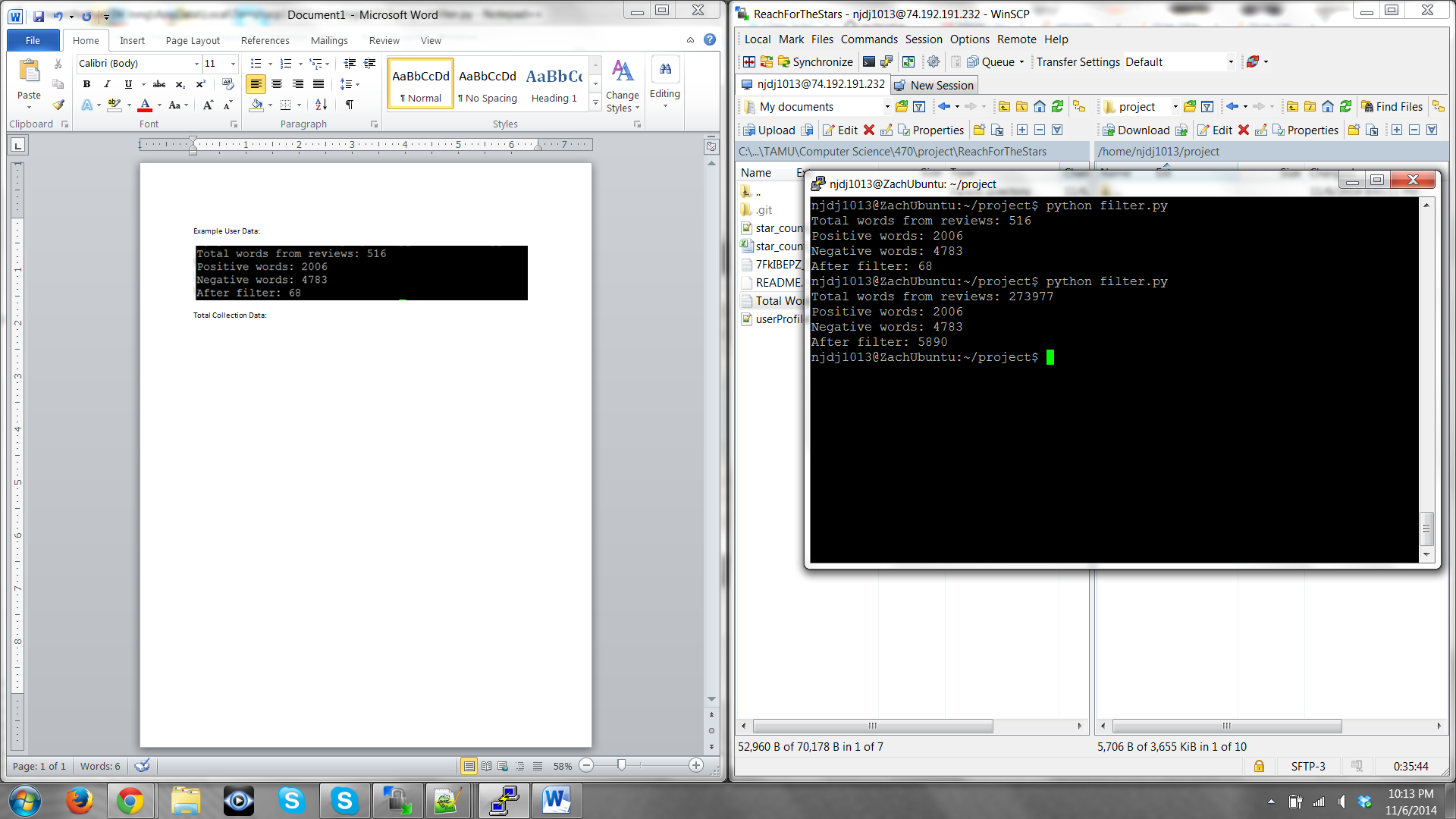
First, we have data for the average rating of collection reviews. We created a map reduce function to simply extract the numeric rating from every review, and to report them. This is useful because it can give us indications of how a specific user rates in relation to the rest of the world. The data is presented below.

Next, we provide two different types of language models – the document collection and an individual user. We can run these through a filter of semantic words to pull out the most meaningful data. Below is a demonstration of this process on the two sets.

User Profile Data:



Total Collection Data:



In an ideal, we create a user profile for every user. However, right now we lack the necessary space and need to support this endeavor. For now we set our targets at a reasonable number (say 100) of user profiles, corresponding to users who have gone to locations with few reviews.