CS432 Assignment #9

1. Care2.com for my blog choice as it had a 100 entries for its RSS feed. I took the most recent blog entries and separated them into 5 categories: Advice, Animals, Ecofriendly, Health, and News. The xml is named “rssfeed.xml.”
2. The first 50 blogs were manually categorized based on what the titles of the entries and the small previews of the entries were and are located in “Manual.txt” within the “data” folder A modified version of a set of scripts made by Kevin Clemmons was used to classify the remaining 50 entries using the fisher classifier. From the train classifier folder, trainClassifier.py was used to produce the results hown in the table.

[This part left intentionally blank.]

