Ell Buscemi Analogy Solver Results

Overall, I was surprised at how smoothly the coding aspect of this project seemed to go, however my program takes a very long time to run (30 minutes to a few hours depending on whether normalization is enabled) and I'm not sure exactly why. The most accurate distance method was usually Euclidean distance, whereas cosine and Manhattan were very close to each other in accuracy. Euclidean distance accordingly took a longer time to run, making the increase in accuracy possibly not worth it. Some types of words were often missing from the model, "family" and "currency" in particular had many analogies that needed to be skipped over due to missing words. There was also a large discrepancy in accuracy between files, for example "gram6-nationality-adjective.txt" had an accuracy of 64.48% with a normalized Manhattan method whereas "gram2-opposite.txt" only had an accuracy of 0.25%. Normalization did not seem to make a difference to accuracy in any case which may be an indication of error (which is a shame because the normalized Euclidean method apparently took about 3 and a half hours to run). I had trouble with the math in this project due to a lack of experience and so I hope that the research sources I found were accurate.