# Computational Journalism: Assignment 7

Pankhuri Kumar (pk2569)

Analysis

One of the most obvious places where OpenCalais fails is recognizing pronouns – it's able to recognize them at the end of quotes, but nowhere else. I'm guessing it is hard-coded into their algorithm that is a quote ends with "he said," it refers to someone who has been previously mentioned.

Secondly, even when OpenCalais reconciles pronouns, it cannot recognize other ways of referring to the same entity - "commander-in-chief" is not the same as "the President" is not the same as "Trump" in the same way that "the company" is not the same as a named company in the article.

OpenCalais is very good at recognizing places, as this is a straight-forward corpus. It counts states, cities and countries separately (while I counted all of it as one) – this is great for reconciling places later if we have a large corpus of documents related to a particular country or region, rather than a specific place.

The other place OpenCalais works quite well is relationships. It is able to recognize that "Bernstein analyst (name of person)" has two different entities – Bernstein and the person. In multiple places, OpenCalais is able to recognize and distinguish these entities, which is great.

There are definitely ambiguities in what is an entity, or how to classify. In certain places, I've marked multiple words together to form an entity, where as OpenCalais separates them out. On the other hand, there are places where I am able to infer that certain pronouns or synonyms refer to the same entity, which OpenCalais is unable to. The specificity of entities is also ambiguous. I'm not sure if the generic term "companies" when refering to a group of companies qualifies as an entity (OpenCalais doesn't think so) while "the company" referring to a specific company is definitely an entity.

Note: The excel document has two sheets: the first one is just a count of the different entities (for my referral), while the second sheet contains the confusion matrices required for the assignment.