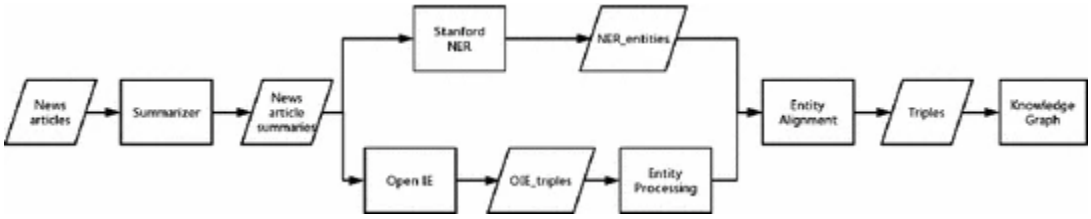
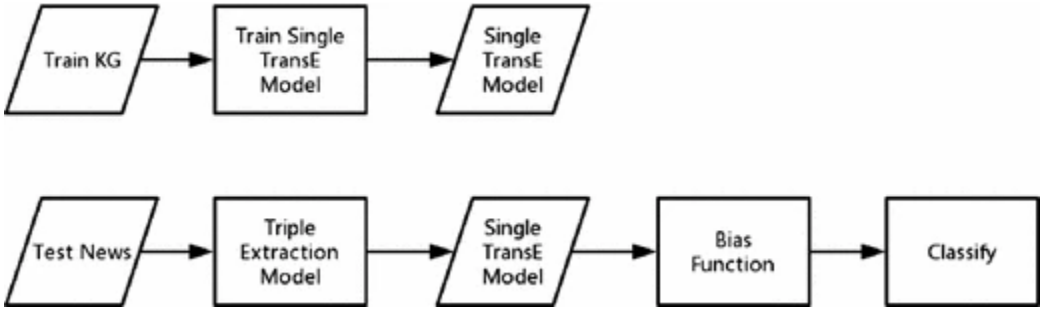
Fact Based - Detection

https://link.springer.com/chapter/10.1007/978-3-030-00671-6_39
Pan, Jeff Z., et al. "Content based fake news detection using knowledge graphs." *The Semantic Web–ISWC 2018: 17th International Semantic Web Conference, Monterey, CA, USA, October 8–12, 2018, Proceedings, Part I 17*. Springer International Publishing, 2018.



This flow chart shows the way this method builds knowledge graphs for fake news detection . It does something interesting called identifying triples. This looks for subject object relationships between known entities that have been identified by the stanford library. They are stored as **(Subject, Predicate, Object)**
For example: *(Elon Musk, CEO of, Tesla) (Apple, headquartered in, Cupertino)*

These triples are then used to create a graph of knowledge that represents what we know about certain individuals. Perhaps we could benefit from this strategy as it seems to be very logically thorough and complete.
Single Model



This next figure shows how the news is then tested up against this knowledge graph and classified mathematically using the bias function. First, a single machine learning model is trained on the knowledge graph, allowing it to place triples into vector space and measure their similarity to other data points. I like this as the triples and bias function provides a strict criteria for how to classify news based on distribution and statistic.

$$f_b(triple_i) = ||\mathbf{h}_i + \mathbf{r}_i - \mathbf{t}_i||_2^2$$

This function represents the way the model determines bias for each triple defined in the model. The classification can then be made in two different ways.

they can be based on the Average bias:

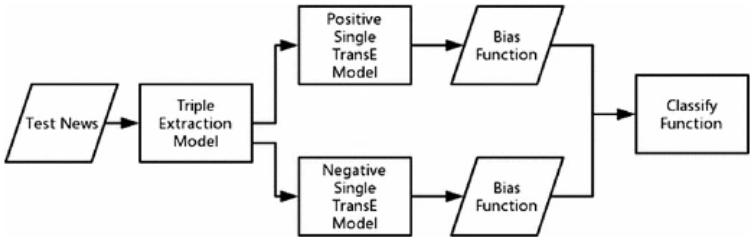
$$f_{avgB}(TS) = \frac{\sum_{i=1}^{n} f_b(triple_i)}{|\mathbf{T}S|}$$

Or the max bias:

$$f_{maxB}(TS) = f_b(triple_{max})$$

I personally find the max bias to be a better option. This is because when it comes to fake news detection, perfection should be the standard. If our fake news detector determines that one fact is extremely biased, this compromises the integrity of the entirety of the source. In this metric, you should only be as strong as your weakest link. I think this idea could be utilized in our project.

Binary Model



Here is an approach where they use both positive and negative models to classify the facts. This is interesting. I like the idea of checking the fact with not only real confirmed news but confirmed fake news. This is necessary because it is not always the case that these models agree on the classification. There are apparently sometimes issues where the news preforms poorly against both models meaning they would have opposite classifications.

> In the B-TransE Model, we represent one news item as $\mathbf{N} = \{\mathbf{TS}, \mathbf{TS}', \mathbf{M}, \mathbf{M}'\}$, $\mathbf{T}S$ refers to triple set extracted from the news based on $\mathbf{M}$ and each triple is defined as $triple_i = (\mathbf{h}_i, \mathbf{t}_i, \mathbf{r}_i)$, and $\mathbf{T}S'$ refers to triple set based on $\mathbf{M}'$ and each triple is $triple'_i = (\mathbf{h}'_i, \mathbf{t}'_i, \mathbf{r}'_i)$, where i refers to the index of the triple in each triple set.

$$f_{mc}(N) = 0, if\, f_b(triple_{max}) < f_b(triple'_{max})$$

$$f_{mc}(N) = 1, otherwise$$

The max function works here by comparing the max bias values of each of the models to determine the classification. 0 meaning true and one being false.

$$f_{ac}(N) = 0, if f_{avgB}(TS) < f_{avgB}(TS')$$

$$f_{ac}(N) = 1, otherwise$$

There is also an average function for this. This paper goes on to propose a hybrid model. Instead of picking one model to base the score on. It creates a vector to hold both max scores for the classification.

> $$V = (Vec_{max}, Vec'_{max})$$