

# Named Entity Disambiguation with Knowledge Graphs

Brian Lin   Shane Ong   Cory Williams   Yun Bin (Matteo) Zhang

## Abstract

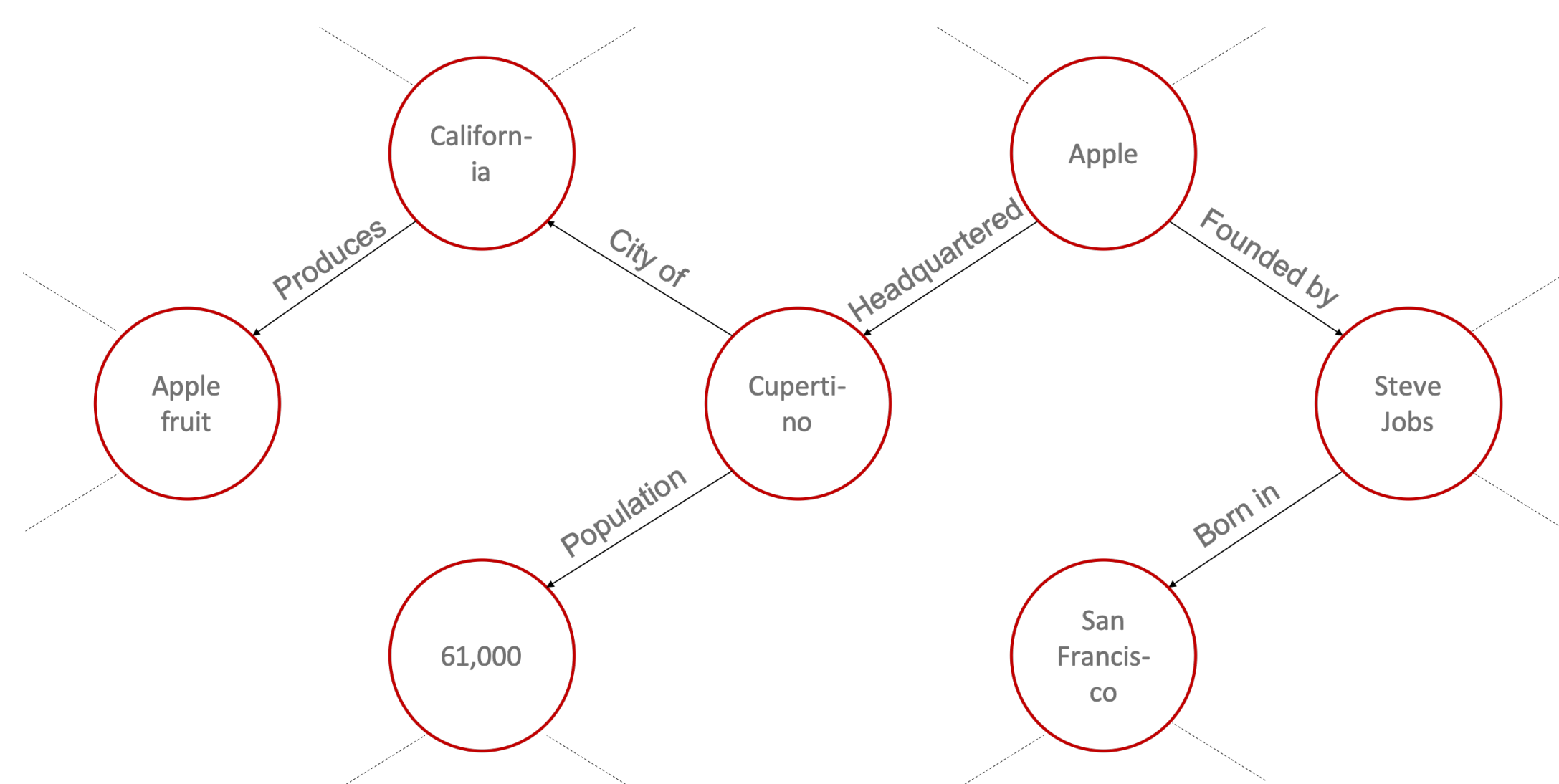
We tackle the Named Entity Disambiguation (NED) problem using a knowledge graph built from Wikidata. Our model uses the spaCy library to perform NER, which identifies about 85% of the Wikidata objects within a text document. Using trained knowledge graph embeddings and text-based contextual embeddings, the model then links each object to the appropriate Wikidata item with an accuracy of 76%.

## Motivation

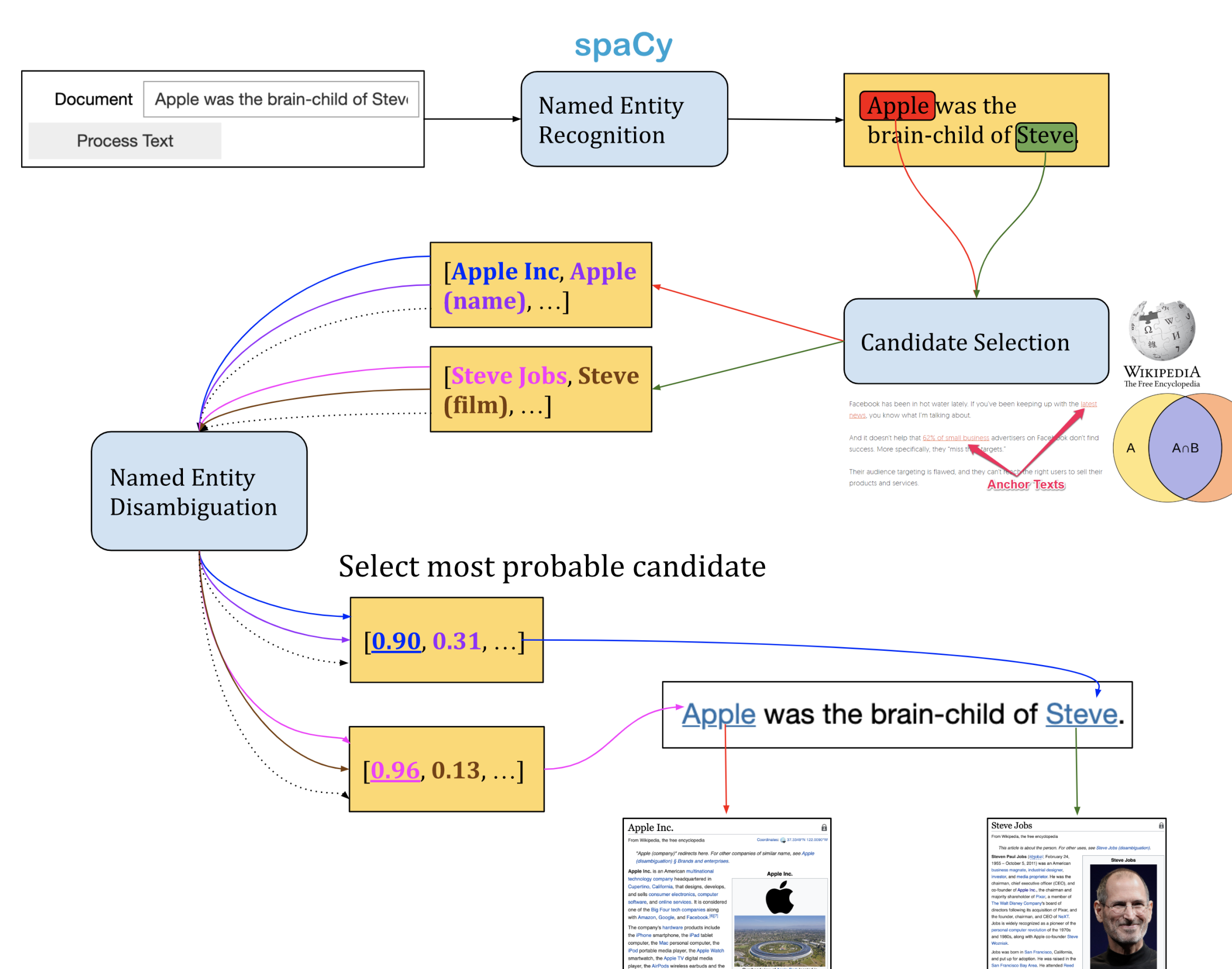
Kensho is interested in using NED to perform text analysis in relation to the financial market. NED is a NLP task which assigns a unique identity to entities mentioned in text. For example, Kensho may want to identify all companies mentioned within a news article, and subsequently investigate how the relations between the companies might affect the markets.

## Knowledge Graphs

Knowledge graphs are a type of database in which entities are represented as nodes, and are connected through edges depicting their relationships. Our knowledge graph was built from Wikidata.

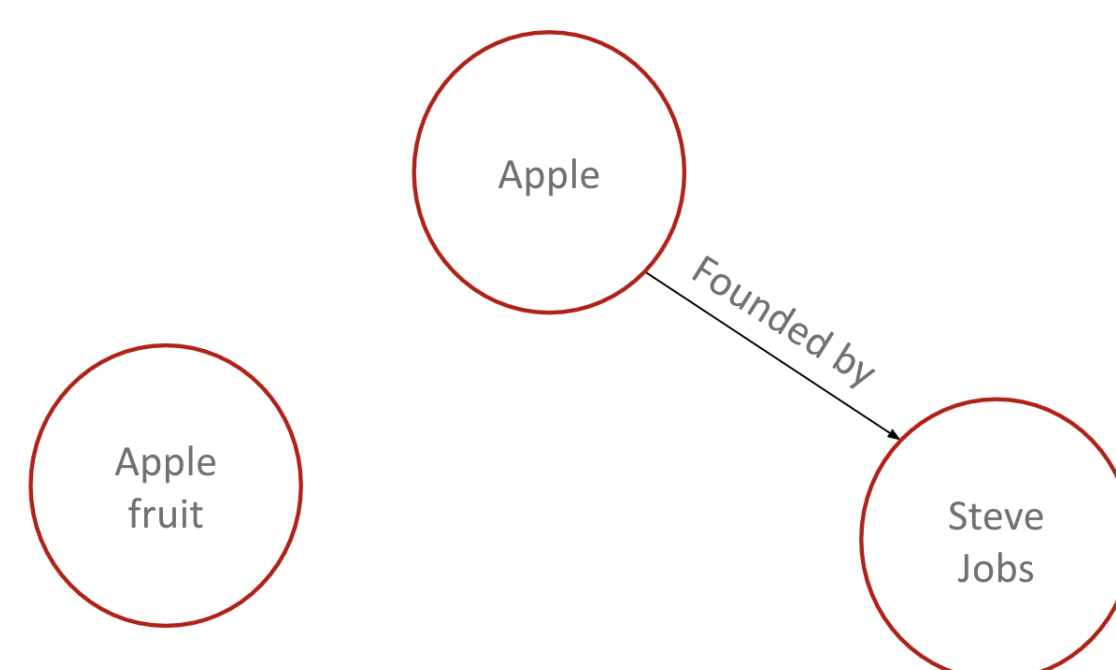


## Overall Pipeline



## Baseline NED Model

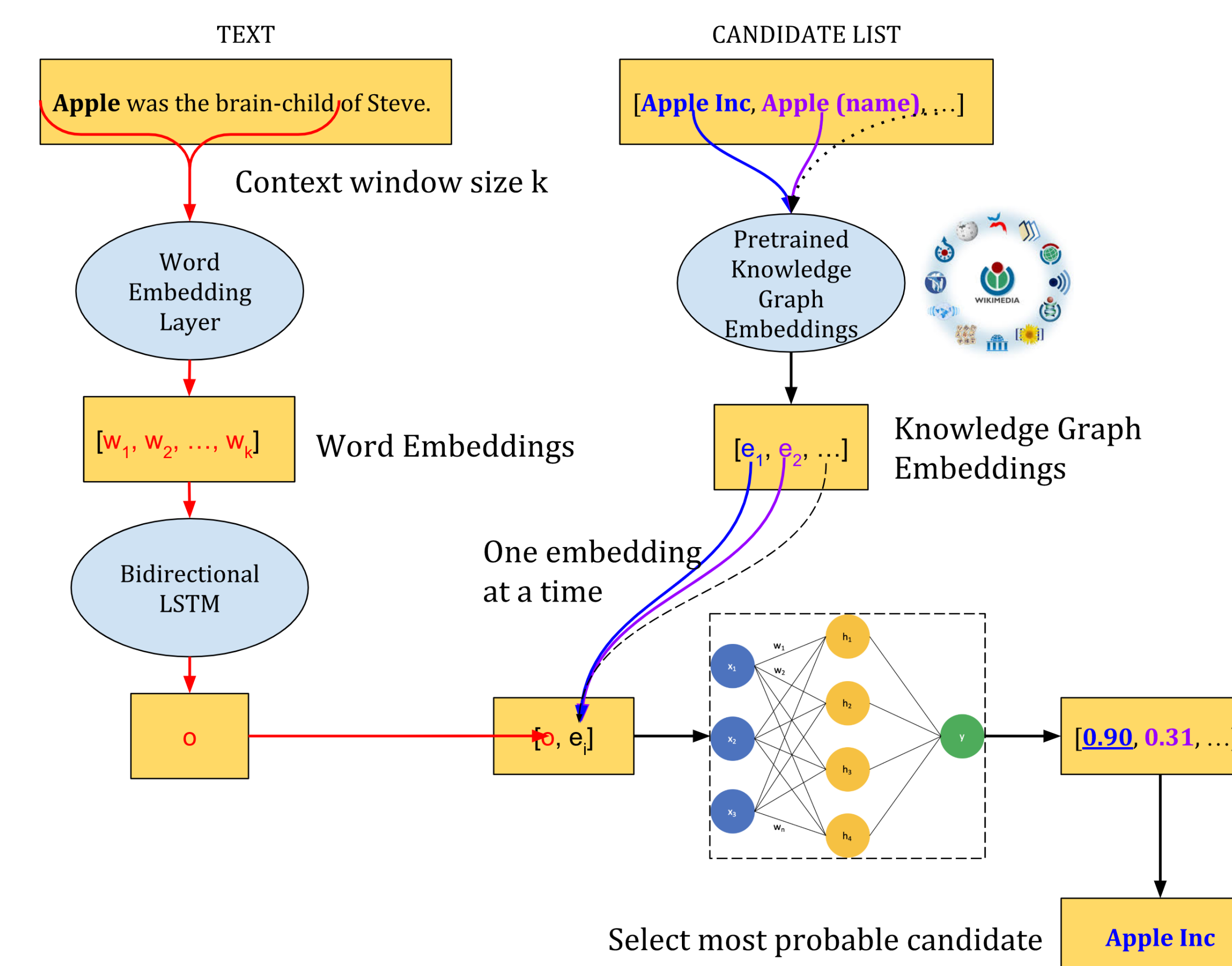
We retrieve direct relations between all the candidate Wikidata items and then choose the candidate with the most connections for each identified entity.



Assuming perfect candidate selection, this model predicts the correct entity 48% of the time. The accuracy is low because we only include nodes directly connected to candidates. It also fails to capture any contextual information in the text.

## Final NED Model

We treat entity linking as a classification problem: We predict whether each candidate is the correct or incorrect entity.



The model predicts the correct entity 76% of the time. This model combines graphical and NLP methods, and outperforms traditional methods.

## Future Work

The simplest extension is to improve the entity recognition and candidate selection stages. Another extension is using joint models boosted with knowledge graphs.

## Acknowledgement

We thank our IACS advisors Pavlos Protopapas, Chris Tanner, and our project partners from Kensho, Georg Kucsco, Gabriel Altay, Aron Szanto, Patrick O'Neill, Mikey Shulman for their advice throughout the semester.