# Mining Evidences for Named Entity Disambiguation in Tweets

Ayesha Siddiqa(CS14S012)
Ditty Mathew(CS13D018)

January 31, 2015

## Problem Statement

Named entity disambiguation is the task of disambiguating named entity mentions in natural language text and link them to the corresponding entries in a knowledge base such as Wikipedia. This process involves identifying named entities from the natural language text(tweets) and mining evidences from external knowledge sources. Mining evidences is the task of finding additional evidences from external sources to improve the named entity disambiguation accuracy.

## Methodology

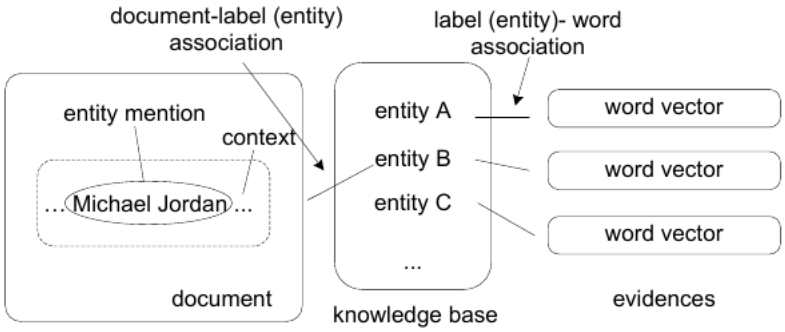


Figure 1: Process Overview

In this work, we are restricting the entity type to *person*. For example, consider the sentence - *Michael Jordan won the best paper award.* Here we have to disambiguate the entity *Michael Jordan* as a scientist rather than as a sports

man. Using the context words( eg: best paper award) around the entity, evidences such as researcher, machine learner, scientist can be mined from external sources. These evidences help us to disambiguate the entity better.

The process overview has shown in Fig 1. This involves document[1] to entity/label association and entity/label to word association.

The proposed project is an extension of [1]. It worked well for tweet data set. But the time complexity is quite large. We are targeting to improve the complexity along with the accuracy.

# References

[1] Fangqiu Han Jiawei Han Dan Roth Xifeng Yan Yang Li, Chi Wang. *Mining Evidences for Named Entity Disambiguation.* KDD 2013.

[1] here we are using tweets as documents