

# Geographical Reference Disambiguation

Xiangrui Meng, Ying Wang

# Missing Metadata

- Motivation: Aid visualization in WDL.
- Identify geographical and chronological references from full text.
- Three steps: recognition, disambiguation, visualization.

# Why We Need to Disambiguate

- Washington – a place or a person?
- Oxford – Oxford, NC or Oxford, UK?



# City Location Disambiguation

- Each city has a list of potential locations. We pick the most probable one.
- Two underlying assumptions:
  - Cities mentioned in a book form several clusters.
  - Consecutive cities in the book should be close to each other.

# City Location Disambiguation

## Our Idea

- Turn the disambiguation problem into an optimization problem!
- In particular, we want to choose locations  $p_1, p_2, \dots, p_n$  for the cities that minimize
$$\text{dist}(p_1, p_2) + \text{dist}(p_2, p_3) + \dots + \text{dist}(p_{n-1}, p_n)$$
- It turns out that this is an optimization problem we can solve!

# Summary

- We present a novel algorithm for disambiguating geographical references.
- The actual accuracy is yet to be determined.