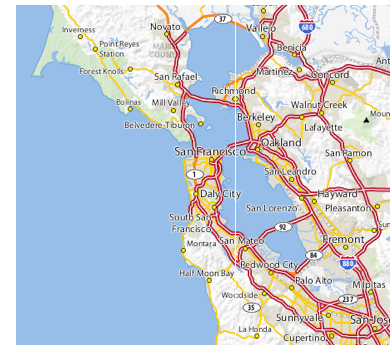


# CS224N Final Project

## Geo-location Route Recognition





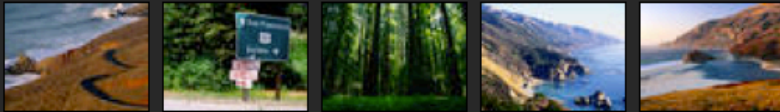
# Objective

---

- We would like that our system can extract a list of locations from web pages that represents the direction of the route and plot the route on a map.



# Example



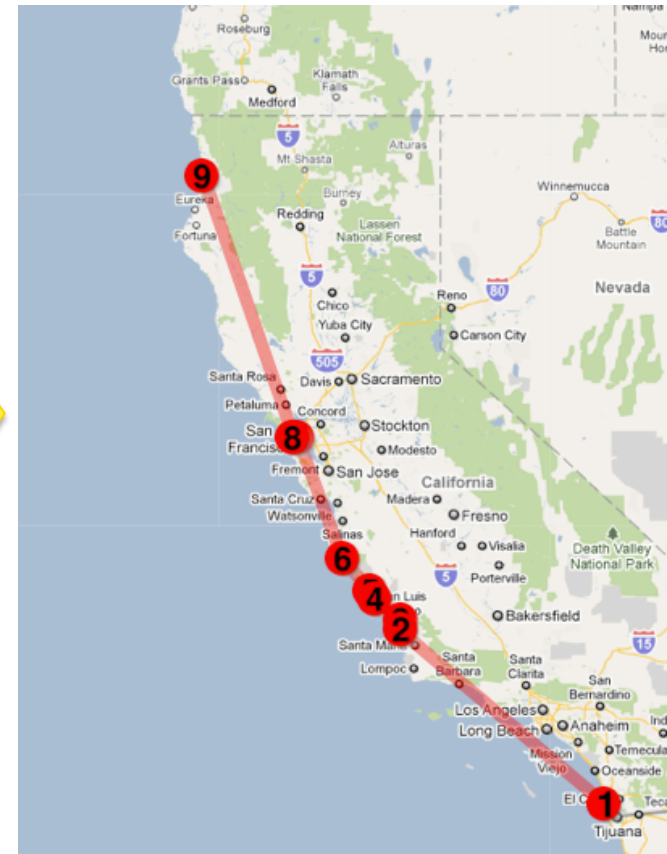
› View gallery

Here's a cherry-picked selection of spots to see along the PCH, an excerpt of the thorough itinerary of the area in our [California Trips](#) book.

Start at the bottom of the state map, just outside [San Diego](#), where the pretty peninsular beach town of [Coronado](#) is connected to the mainland by a long, narrow spit of sand called the Silver Strand. You can't help but feel a thrill while speeding over the 2-mile-long Coronado Bridge as it snakes its way across San Diego Bay.

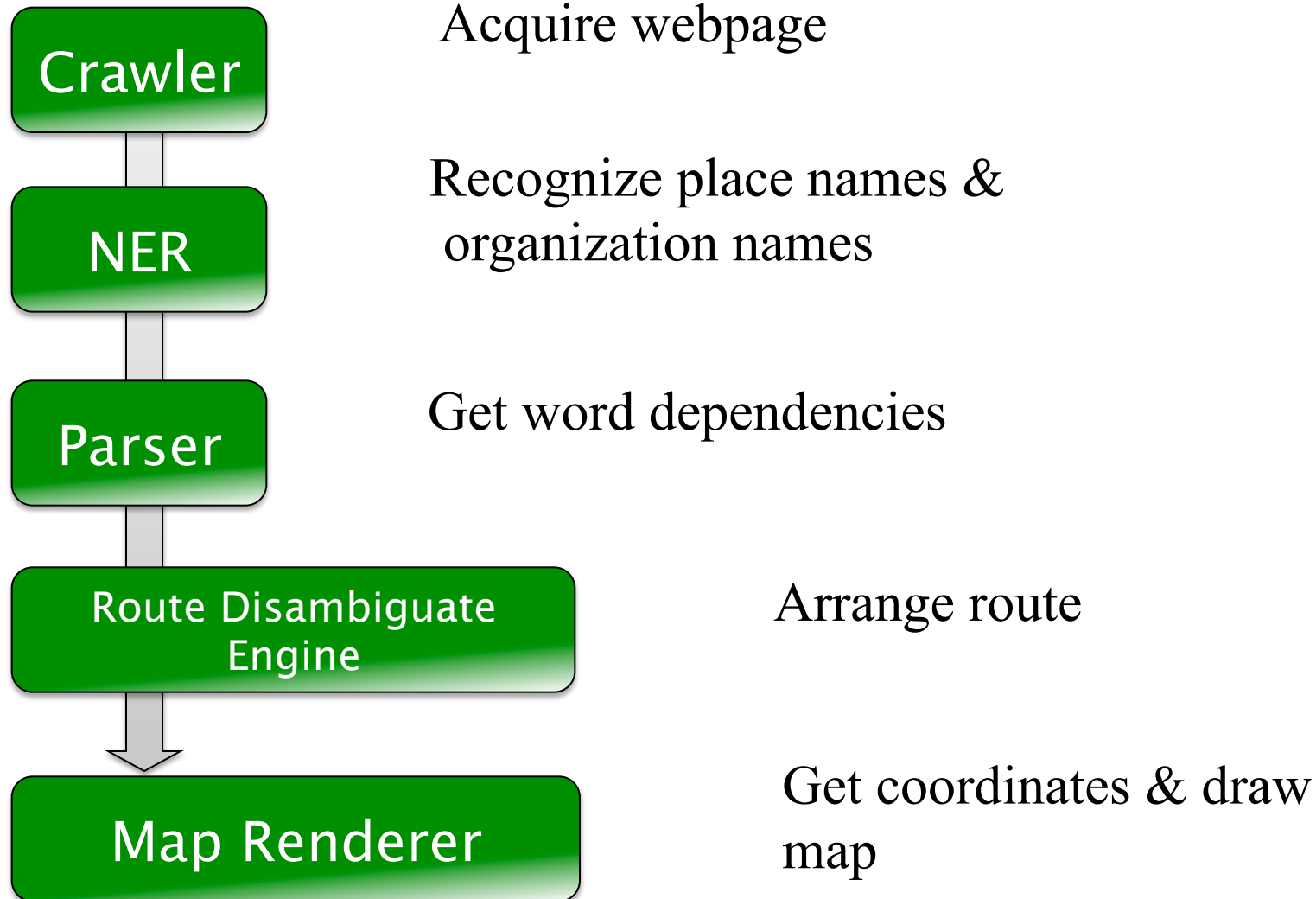
Further along the track, head down toward the sea and [Torrey Pines State Natural Reserve](#), harboring the rarest pines in North America and a lagoon frequented by migrating seabirds. This stretch of coast is astonishingly wild. On to Long Beach and the often overlooked [Long Beach Museum of Art](#), sitting pretty on an ocean bluff. Exhibits focus on California modernism, with surprisingly cool contemporary art. In an early-20th-century mansion, the museum's gourmand café has killer weekend brunches and postcard-perfect views.

*From [www.lonelyplanet.com](http://www.lonelyplanet.com)*





# Pipeline

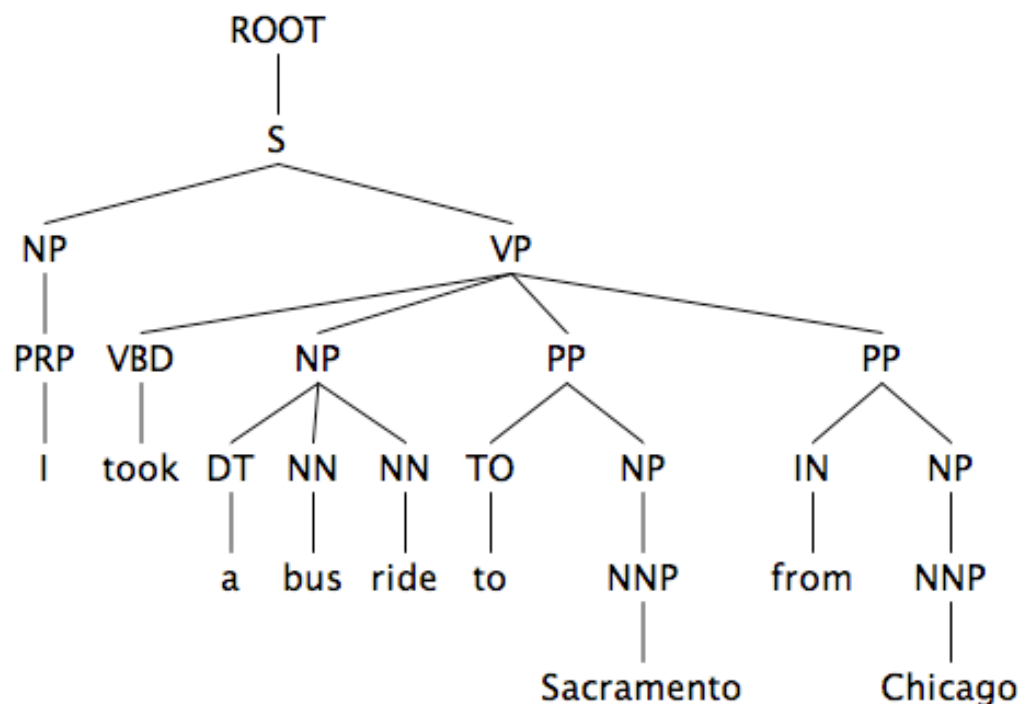




# From Typed Dependency to Route

## *Prepositional Phrase*

- I took a bus ride to Sacramento from Chicago.



nsubj(took-2, I-1)

det(ride-5, a-3)

nn(ride-5, bus-4)

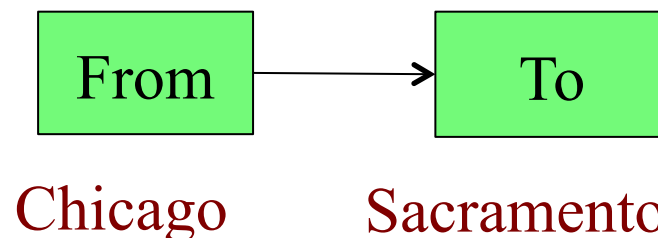
dobj(took-2, ride-5)

prep(took-2, to-6)

**pobj(to-6, Sacramento-7)**

prep(took-2, from-8)

**pobj(from-8, Chicago-9)**





# Evaluation

- **Score** = 
$$\frac{\text{locations in the golden test data}}{\text{locations in the golden test data} + \text{edit distance}}$$
- **Precision**: We generate lists of unique places appearing in the test program output and the golden test data separately, then match them to find out how many locations appear in both, then calculated precision using the matching and total line counts.
- **Recall**: We calculate recall by dividing the matching lines by the total lines in the golden test data.



# Test and Results

- Data
  - Forum data from [www.lonelyplanet.com](http://www.lonelyplanet.com)
- Baseline
  - Start and end point according to the order of appearance
- Method
  - Look five sentences in a forum page
- Result

	Precision	Recall	Score
Our system	0.549	0.602	0.438
Baseline	0.537	0.454	0.588



# Example Output

Best way to get from **San Cristobal de las Casas** to **San Miguel de Allende**?

Hello, I need to travel from **San Cristobal de las Casas**, in Chiapas, to **San Miguel de Allende**, in Guanajuato, as inexpensively and easily as possible.

I took the ADO overnight bus from **Oaxaca** to **San Cristobal** for 488 pesos.

I am shocked to find that the bus from **San Cristobal** to **Mexico City** is nearly 1000 pesos, plus then I have to get to **San Miguel**, another 4-5 hour bus ride, and haven't found prices yet.





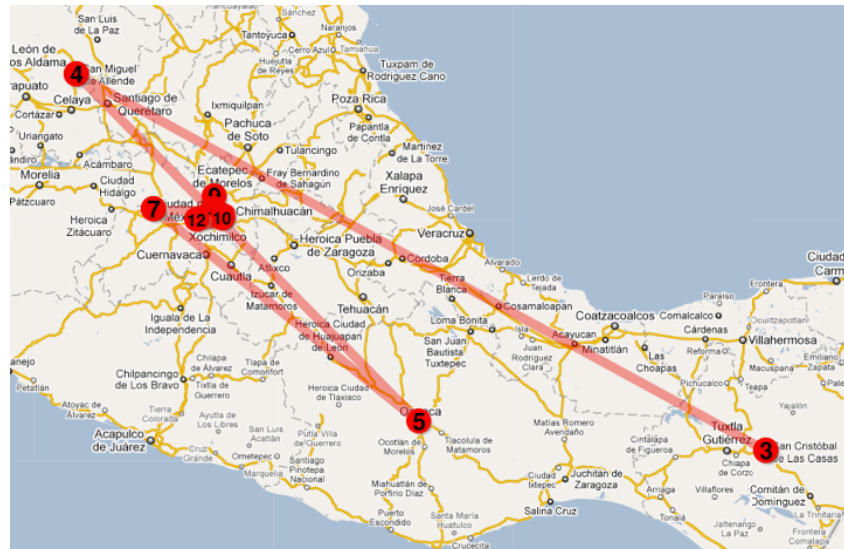
# Example Output Locations

## Output Route

San Cristobal de las Casas  
San Miguel de Allende  
San Cristobal de las Casas  
San Miguel de Allende  
Oaxaca  
San Cristobal  
San Cristobal  
Mexico City  
San Miguel

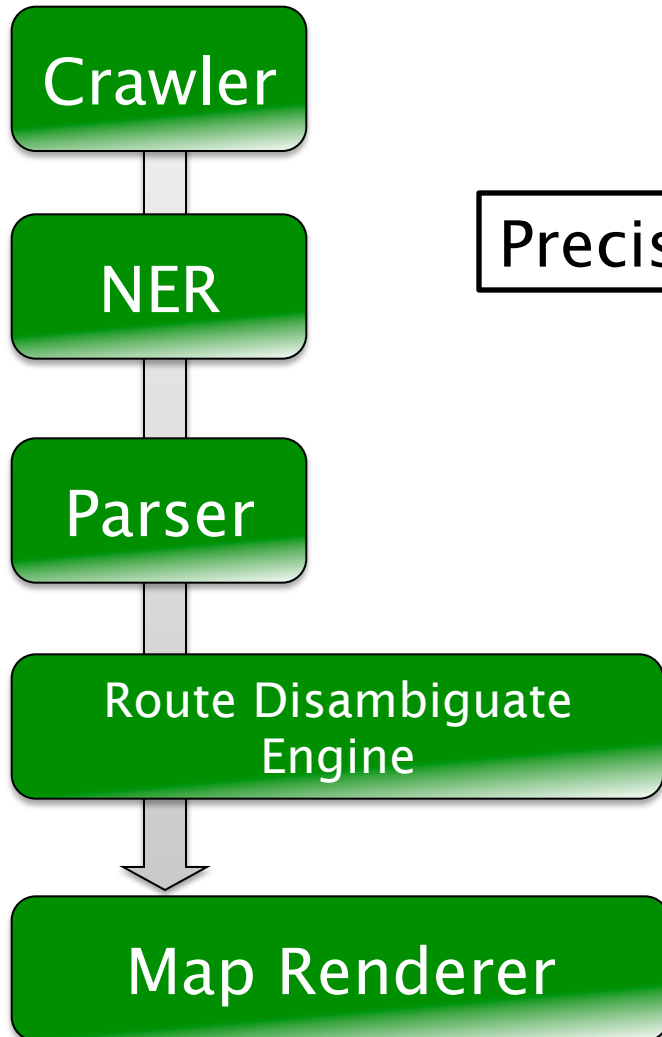
## Golden Route

San Cristobal de las Casas  
Tuxla Gutierrez  
Mexico City  
San Miguel de Allende





# Problems and Future Work



Precision and Recall of the NER system

How to recognize different routes in one document according to context

Location ambiguity  
Cambridge:  
Cambridge, MA or  
Cambridge, UK