CSE 454
Advanced Internet Systems

Project Logistics

Dan Weld

Schedule

• Teams
  – Today
• Bidding
  – Starting tonight
• Meetings
  – Thurs / Fri

Challenges

• Unfortunately for my slot, there is only one positive result in the test set, so I can't be very sure how well I am doing.

Are Slot Values Really Rare?

• Gold-labeled data
• Test Data
• Actual NIST Challenge 2010
  – Training data
  – Queries “Joe Biden”
  – Union of all system’s output (just on these queries)
  – Human checking
• 2011, 2012

Challenges

• Raul Castro was born on June 3, 1931 in Cuba ’s Oriente Province and educated at Jesuit schools and the University of Havana.
  – My pattern catches *Oriente* but not *Oriente Province*.
• President Hugo Chavez was born here, his father has governed it for 10 years, and now his brother is seeking the governor’s job.

Challenges

• Hugo Chavez, was born on July 28, 1954, in Venezuela ’s Sabaneta.
  – Because any amount of text can be in-between *born* and *in Venezuela ’s Sabaneta*, it is hard to know if *in* is referring to *born*.
• Hugo Chavez was born before skinny dipping at a swimming pool in Venezuela ’s Sabaneta.
In 1953, five years after the state was established the JNF was dissolved and re-organized as an Israeli company.

- This should have been paired with the Jewish National Fund.

My final solution, which is pretty terrible, doesn’t use any regular expressions. It first eliminates totally incorrect answers by skipping ignored-slots and entities (non-“ORG” types). Then it checks if there’s a date and an organization...Then it just checks if the word ‘dissolved’ appears anywhere in the sentence. If so, it links the org to the date.

Recall: 1/3 = 0.33333334
Precision: 1/50 = 0.02
F1: 0.03735858