NLP
Introduction to NLP

Relation extraction
Relation extraction

• Links between entities
  – Works–for
  – Manufactures
  – Located–at
Relation Extraction

- **Person–person**
  - ParentOf, MarriedTo, Manages
- **Person–organization**
  - WorksFor
- **Organization–organization**
  - IsPartOf
- **Organization–location**
  - IsHeadquarteredAt
ACE Evaluation

• 2002 newspaper data
• Entities:
  – Person, Organization, Facility, Location, Geopolitical Entity
• Relations:
  – Role, Part, Located, Near, Social
Relation Extraction

• Core NLP task
  – Used for building knowledge bases, question answering

• Input
  – Mazda North American Operations *is headquartered in Irvine, Calif.*, and oversees the sales, marketing, parts and customer service support of Mazda vehicles in the United States and Mexico through nearly 700 dealers.

• Output
  – IsHeadquarteredIn (Mazda North American Operations, Irvine)
Relation extraction

• Using patterns
  – Regular expressions
  – Gazetteers

• Supervised learning

• Semi-supervised learning
  – Using seeds
Extracting IS-A Relations

• Hearst’s patterns
  – X and other Y
  – X or other Y
  – Y such as X
  – Y, including X
  – Y, especially X

• Example
  – Evolutionary relationships between the platypus and other mammals
Supervised Relation Extraction

• Look for sentences that have two entities that we know are part of the target relation
• Look at the other words in the sentence, especially the ones between the two entities
• Use a classifier to determine whether the relation exists
Example

• English
  – Beethoven was born in December 1770 in Bonn
  – Born in Bonn in 1770, Beethoven ...
  – After his birth on December 16, 1770, Beethoven grew up in a musical family
  – Ludwig van Beethoven (1770–1827)
  – While this evidence supports the case for 16 December 1770 as Beethoven's date of birth
Example (non-English)

- **German**
  - Ludwig van Beethoven *wurde* am 17. Dezember 1770 in Bonn *getauft*
  - Ludwig van Beethoven *wurde* in Bonn, 15. Dezember 1770, eine Familie ursprünglich aus Brabant in Belgien *geboren*
  - Der Geburtstag von Ludwig van Beethoven wurde im Winter 1770 in Bonn nicht genau dokumentiert

- **Spanish**
  - Ludwig van Beethoven *nació* en Bonn el 17 de diciembre de 1770
  - *Nacido* en Bonn 1770, Beethoven ...
  - Ludwig van Beethoven, *nace* en diciembre de 1770
Semi-supervised Relation Extraction

- Start with some seeds, e.g.,
  - Beethoven was born in December 1770 in Bonn
- Look for other sentences with the same words
- Look for expressions that appear nearby
- Look for other sentences with the same expressions
Evaluating Relation Extraction

- **Precision** $P$
  - correctly extracted relations/all extracted relations
- **Recall** $R$
  - correctly extracted relations/all existing relations
- **F1 measure**
  - $F1 = \frac{2PR}{P+R}$
- **If there is no annotated data**
  - only measure precision
NLP