

# Relation Extraction

# What is relation extraction?



## Extracting relations from text

- Company report: “International Business Machines Corporation (IBM or the company) was incorporated in the State of New York on June 16, 1911, as the Computing-Tabulating-Recording Co. (C-T-R)...”

- Extracted Complex Relation:

### Company-Founding

Company	IBM
Location	New York
Date	June 16, 1911
Original-Name	Computing-Tabulating-Recording Co.

- But we will focus on the simpler task of extracting relation **triples**

Founding-year(IBM,1911)

Founding-location(IBM,New York)



# WIKIPEDIA

The Free Encyclopedia

Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia

Interaction
Help
About Wikipedia
Community portal
Recent changes
Contact Wikipedia

Toolbox
Print/export

Languages
العربية
Azerbaycanca
Беларуская
Беларуская (тарашкевіца)

English blows! FORD FOUNDER



Stanford EQ Leland Stanford Junior University  
 Stanford LOC IN California  
 Stanford IS A research university  
 Turing award Motto Die Luft der Freiheit weht  
 Stanford LOC NEAR Palo Alto (German)<sup>(1)</sup>  
 Stanford FOUNDED IN 1891  
 Motto in English The wind of freedom blows<sup>(1)</sup>  
 Humanities Stanford FOUNDER Leland Stanford



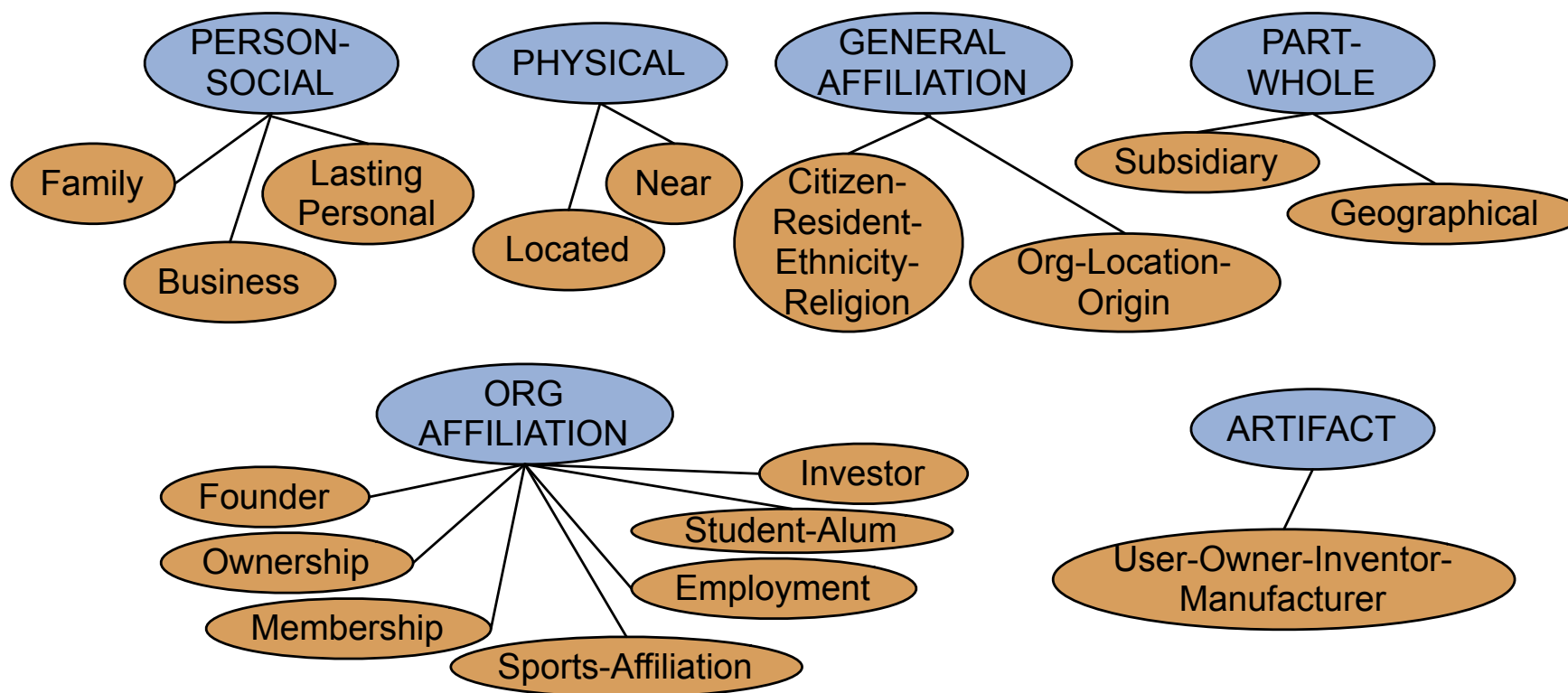
## Why Relation Extraction?

- Create new structured knowledge bases, useful for any app
- Augment current knowledge bases
  - Adding words to WordNet thesaurus, facts to FreeBase or DBPedia
- Support question answering
  - The granddaughter of which actor starred in the movie “E.T.”?  
(acted-in ?x “E.T.”)(is-a ?y actor)(granddaughter-of ?x ?y)
- But which relations should we extract?



# Automated Content Extraction (ACE)

17 relations from 2008 "Relation Extraction Task"





# Automated Content Extraction (ACE)

- Physical-Located **PER-GPE**  
`He` was in `Tennessee`
- Part-Whole-Subsidiary **ORG-ORG**  
`XYZ`, the parent company of `ABC`
- Person-Social-Family **PER-PER**  
`John's` wife `Yoko`
- Org-AFF-Founder **PER-ORG**  
`Steve Jobs`, co-founder of `Apple...`

Dan Jurafsky



# UMLS: Unified Medical Language System

- 134 entity types, 54 relations

Injury	disrupts	Physiological Function
Bodily Location	location-of	Biologic Function
Anatomical Structure	part-of	Organism
Pharmacologic Substance	causes	Pathological Function
Pharmacologic Substance	treats	Pathologic Function



## Extracting UMLS relations from a sentence

Doppler echocardiography can be used to  
diagnose left anterior descending artery  
stenosis in patients with type 2 diabetes



Echocardiography, Doppler **DIAGNOSES** Acquired stenosis





# Databases of Wikipedia Relations

## Wikipedia Infobox

Relations extracted from Infobox

Stanford **state** California

Stanford **motto** "Die Luft der Freiheit weht"

```
{{Infobox university
|image_name= Stanford University seal.svg
|image_size= 210px
|caption = Seal of Stanford University
|name =Stanford University
|native_name =Leland Stanford Junior Uni
|motto = {{lang|de|"Die Luft der Freiheit v
name="casper">{{cite speech|title=Die Lu
Casper|first=Gerhard|last=Casper|author
05|url=http://www.stanford.edu/dept/pr
|mottoeng = The wind of freedom blows<
|established = 1891<ref>{{cite web |
url=http://www.stanford.edu/home/stan
publisher = Stanford University | accessd
|type = [[private university|Private]]
|calendar= Quarter
|president = [[John L. Hennessy]]
|provost = [[John Etchemendy]]
|city = [[Stanford, California|Stanford]]
|state = California
|country = U.S.
```

Type	Private
<b>Endowment</b>	US\$ 16.5 billion (2011) <sup>[3]</sup>
<b>President</b>	John L. Hennessy
<b>Provost</b>	John Etchemendy
<b>Academic staff</b>	1,910 <sup>[4]</sup>
<b>Students</b>	15,319
<b>Undergraduates</b>	6,878 <sup>[5]</sup>
<b>Postgraduates</b>	8,441 <sup>[5]</sup>
<b>Location</b>	Stanford, California, U.S.
<b>Campus</b>	Suburban, 8,180 acres (3,310 ha) <sup>[6]</sup>
<b>Colors</b>	Cardinal red and white

d  
tml}}</ref>

ty History |



# Relation databases that draw from Wikipedia

- Resource Description Framework (RDF) triples  
subject predicate object  
Golden Gate Park [location](#) San Francisco  
dbpedia:Golden\_Gate\_Park [dbpedia-owl:location](#) dbpedia:San\_Francisco
- DBPedia: 1 billion RDF triples, 385 from English Wikipedia
- Frequent Freebase relations:

people/person/nationality,	location/location/contains
people/person/profession,	people/person/place-of-birth
biology/organism_higher_classification	film/film/genre



# Ontological relations

Examples from the WordNet Thesaurus

- IS-A (hypernym): subsumption between classes
  - Giraffe IS-A ruminant IS-A ungulate IS-A mammal IS-A vertebrate IS-A animal...
- Instance-of: relation between individual and class
  - San Francisco instance-of city



# How to build relation extractors

1. Hand-written patterns
2. Supervised machine learning
3. Semi-supervised and unsupervised
  - Bootstrapping (using seeds)
  - Distant supervision
  - Unsupervised learning from the web



# What is relation extraction?