# **Relation extraction**

Bill MacCartney CS224u Stanford University

Overview

### **Overview**

- The task of relation extraction
- Data resources
- Problem formulation
- Evaluation
- Simple baselines
- Directions to explore

- Task definition
- Goal: machine reading
- Practical applications
- Hand-built patterns
- Supervised learning
- Distant supervision

# The task of relation extraction **Task definition**

If you want to see one in action, Robert Scoble got a ride in the first production model, driven by Tesla Motors chairman Elon Musk.

relation	subject	object
founders	PayPal	Elon_Musk
founders	SpaceX	Elon_Musk
has_spouse	Elon_Musk	Talulah_Riley
worked at	Elon Musk	Tesla Motors

### The task of relation extraction Goal: machine reading

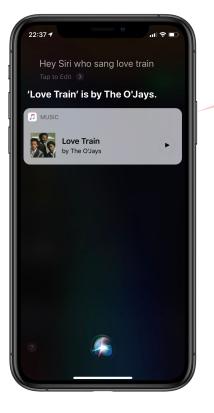
#### Reading the Web: A Breakthrough Goal for AI

I believe AI has an opportunity to achieve a true breakthrough over the coming decade by at last solving the problem of reading natural language text to extract its factual content. In fact, I hereby offer to bet anyone a lobster dinner that by 2015 we will have a computer program capable of automatically reading at least 80% of the factual content [on the] web, and placing those facts in a structured knowledge base. The significance of this AI achievement would be tremendous: it would immediately increase by many orders of magnitude the volume, breadth, and depth of ground facts and general knowledge accessible to knowledge based AI programs. In essence, computers would be harvesting in structured form the huge volume of knowledge that millions of humans are entering daily on the web in the form of unstructured text.



— Tom Mitchell, 2005

## Applications: intelligent assistants



/music/artist/	track	
The O'Jays	Love	Train

Cardi B Bodak Yellow Selena Gomez Bad Liar

#### /film/film/starring

Wonder Woman	Gal Gadot
Dunkirk	Tom Hardy
Tomb Raider	Alicia Vikander

#### /organization/organization/parent

Facebook
Google
Microsoft

#### /people/person/date\_of\_death

 Barbara Bush
 2018-04-17

 Milos Forman
 2018-04-14

 Winnie Mandela
 2018-04-11

"Love Train" is a hit single by The O'Jays, written by Kenny Gamble and Leon Huff. Released in 1972, it reached number one on both the R&B Singles and the Billboard Hot 100, in February and March 1973 respectively, number 9 on the UK Singles Chart and was certified gold by the RIAA. It was The O'Jays' first and only number-one record on the US pop chart.

## Applications: building ontologies

video game action game ball and paddle game Breakout. platform game Donkey Kong shooter arcade shooter Space Invaders first-person shooter Call of Duty third-person shooter Tomb Raider adventure game text adventure graphic adventure strategy game 4X game Civilization tower defense Plants vs. Zombies



----

Mirror ran a headline questioning whether the killer's actions were a result of playing Call of Duty, a first-person shooter game ...

Melee, in video game terms, is a style of elbow-drop hand-to-hand combat popular in first-person shooters and other shooters.



Tower defense is a kind of real-time strategy game in which the goal is to protect an area or place and prevent enemies from reaching ...

## Applications: gene regulation

	Patrone	of he publication. April 10, 2004, and to control from. April 12, 200 a AMI Pageon in Price, 2010 (20, 2000) 102 11 (21 April 2004)2010
Alteriorem Alteriorem Alteriorem Anteri		An exist of the constraint of the distribution
The different manufactor of balance With the second press	tenti ne 1788, primi terapat biane 148188. Anno Arlantarene, 1888, henno anterne mendecatalanen, 1869, tama terapat bian rena, 1852, metter mayor quine 1847, 1	in disc imported by per obtaining in the despinate with which immeries removerables of its sample and sequence from per- teriority to momentality (2), 20, 20, 20, 20, 20, 20, 20, 20, 20, 20
Support and	rine at high more the ang	500

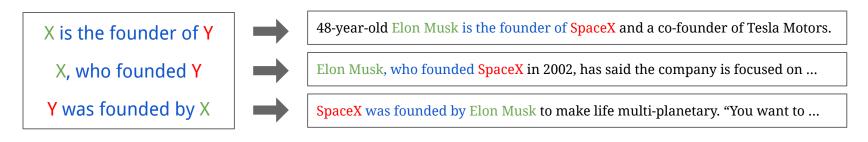
summary for human

relation	subject	object
is_a	p53	protein
is_a	Bax	protein
has_function	p53	apoptosis
has_function	Bax	induction
involved_in	apoptosis	cell_death
is_in	Bax	cytoplasm
related_to	apoptosis	caspase_activation

structured knowledge extraction: summary for machine

# The task of relation extraction Hand-built patterns

### Idea: define some extraction patterns



### Problem: most occurrences do not fit simple patterns

You may also be thinking of Elon Musk (founder of SpaceX), who started PayPal.

Elon Musk, co-founder of PayPal, went on to establish SpaceX, one of the most ...

If Space Exploration (SpaceX), founded by Paypal pioneer Elon Musk succeeds, ...

# The task of relation extraction Supervised learning

### Idea: label examples, train a classifier

You may also be thinking of Elon Musk (founder of SpaceX), who started PayPal.	
Elon Musk, co-founder of PayPal, went on to establish SpaceX, one of the most	
If Space Exploration (SpaceX), founded by Paypal pioneer Elon Musk succeeds,	
Entrepreneur Elon Musk announced the latest addition to the SpaceX arsenal	
Elon Musk tweeted Friday that SpaceX employees are "working on ventilators"	

founder	+1.56
founded	+1.41
establish	+1.23
co-founder	+1.01
PayPal	+0.35
announced	-0.23
addition	-0.32
tweeted	-0.44

### Success! Better generalizability

Problem: labeling examples is expensive :-(

# The task of relation extraction **Distant supervision**

Idea: derive labels from an existing knowledge base (KB)

Assume sentences with related entities are positive examples

Assume sentences with unrelated entities are negative examples

relation	subject	object
founders	SpaceX	Elon_Musk
founders	Apple	Steve_Jobs
founders	Microsoft	Bill_Gates

Elon Musk, co-founder of PayPal, went on to establish SpaceX, one of the most ...

Entrepreneur Elon Musk announced the latest addition to the SpaceX arsenal ...

Elon Musk dismissed concerns that Apple was poaching the company's talent.

Now we know what Apple would have done with Elon Musk if that deal had ...

Hooray! Massive quantities of training data, practically free! Qualm: are those assumptions reliable?





## Distant supervision: limitations

Distant supervision is a powerful idea — but it has two limitations:

1. Not all sentences with related entities are truly positive examples

Entrepreneur Elon Musk announced the latest addition to the SpaceX arsenal ...



(but the benefit of *more* data outweighs the harm of noisier data)

2. Need an existing KB to start from — can't start from scratch