
Hasso Plattner Institut

IT Systems Engineering | Universität Potsdam

# Natural Language Processing

*Relation Extraction*

Potsdam, 14 June 2012

**Saeedeh Momtazi**
Information Systems Group

based on the slides of the course book

# Outline



1. Introduction
2. Task
3. Pattern Extraction
4. Supervised Learning
5. Semi-supervised Learning

# Outline

# Information Extraction



- Named entity recognition
- Relation Extraction

# Named Entity Recognition



HPI is affiliated to the Potsdam University and located in Potsdam near Berlin. It was founded in 1998 by Hasso Plattner, one of the co-founders of the European software company, SAP AG.

.

| | |
|---|---|
| ORG | HPI |
| ORG | Potsdam University |
| LOC | Potsdam |
| LOC | Berlin |
| DATE | 1998 |
| PER | Hasso Plattner |
| ORG | SAP AG |

# Relation Extraction



HPI is affiliated to the Potsdam University and located in Potsdam near Berlin. It was founded in 1998 by Hasso Plattner, one of the co-founders of the European software company, SAP AG.

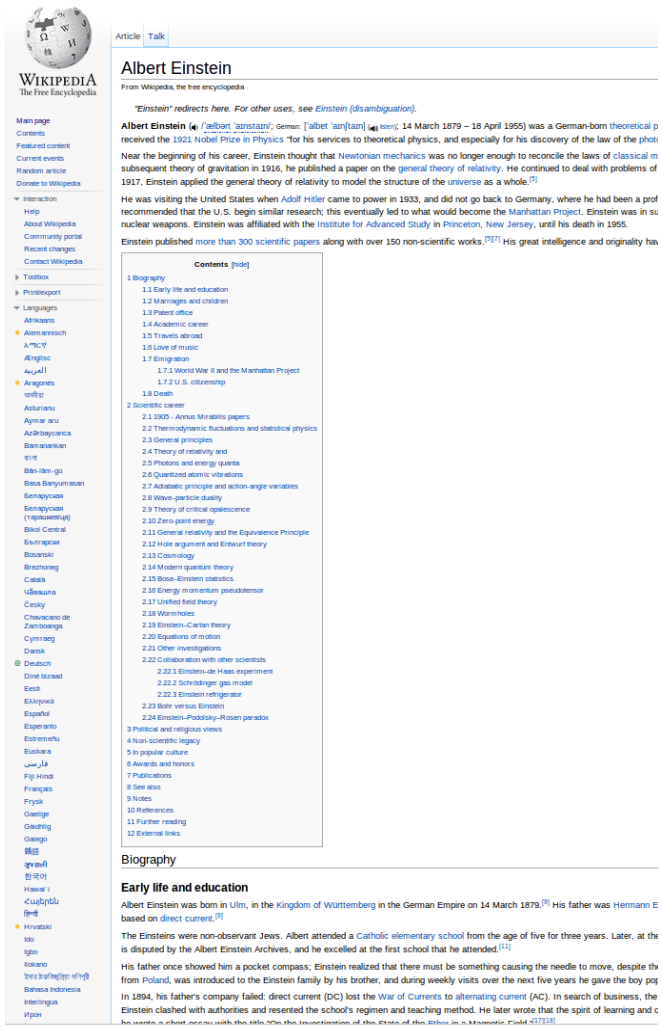
| | |
|---|---|
| located (ORG-LOC) | HPI - Potsdam |
| near (ORG-LOC) | HPI - Berlin |
| near (LOC-LOC) | Potsdam - Berlin |
| founded (ORG-DATE) | HPI - 1998 |
| founder (ORG-PER) | HPI - Hasso Plattner |
| co-founder (ORG-PER) | SAP AG - Hasso Plattner |

# Information Extraction



# Information Extraction





Albert Einstein

From Wikipedia, the free encyclopedia

*"Einstein" redirects here. For other uses, see Einstein (disambiguation).*

**Albert Einstein** (/ˈælbərt ˈaɪnstaɪn/; German: [ˈalbɛʁt ˈaɪnʃtaɪn] (listen); 14 March 1879 – 18 April 1955) was a German-born theoretical p received the 1921 Nobel Prize in Physics "for his services to theoretical physics, and especially for his discovery of the law of the phot

Near the beginning of his career, Einstein thought that Newtonian mechanics was no longer enough to reconcile the laws of classical m subsequent theory of gravitation in 1916, he published a paper on the general theory of relativity. He continued to deal with problems of 1917, Einstein applied the general theory of relativity to model the structure of the universe as a whole.[5]

He was visiting the United States when Adolf Hitler came to power in 1933, and did not go back to Germany, where he had been a prof recommended that the U.S. begin similar research; this eventually led to what would become the Manhattan Project. Einstein was in su nuclear weapons. Einstein was affiliated with the Institute for Advanced Study in Princeton, New Jersey, until his death in 1955.

Einstein published more than 300 scientific papers along with over 150 non-scientific works.[5][7] His great intelligence and originality ha

Contents [hide]

1 Biography
- 1.1 Early life and education
- 1.2 Marriages and children
- 1.3 Patent office
- 1.4 Academic career
- 1.5 Travels abroad
- 1.6 Love of music
- 1.7 Emigration
  - 1.7.1 World War II and the Manhattan Project
  - 1.7.2 U.S. citizenship
- 1.8 Death

2 Scientific career
- 2.1 1905 - Annus Mirabilis papers
- 2.2 Thermodynamic fluctuations and statistical physics
- 2.3 General principles
- 2.4 Theory of relativity and
- 2.5 Photons and energy quanta
- 2.6 Quantized atomic vibrations
- 2.7 Adiabatic principle and action-angle variables
- 2.8 Wave–particle duality
- 2.9 Theory of critical opalescence
- 2.10 Zero-point energy
- 2.11 General relativity and the Equivalence Principle
- 2.12 Hole argument and Einstein's theory
- 2.13 Cosmology
- 2.14 Modern quantum theory
- 2.15 Bose–Einstein statistics
- 2.16 Energy momentum pseudotensor
- 2.17 Unified field theory
- 2.18 Wormholes
- 2.19 Einstein–Cartan theory
- 2.20 Equations of motion
- 2.21 Other investigations
- 2.22 Collaboration with other scientists
  - 2.22.1 Einstein–de Haas experiment
  - 2.22.2 Schrödinger gas model
  - 2.22.3 Einstein refrigerator
- 2.23 Bohr versus Einstein
- 2.24 Einstein–Podolsky–Rosen paradox

3 Political and religious views
4 Non-scientific legacy
5 In popular culture
6 Awards and honors
7 Publications
8 See also
9 Notes
10 References
11 Further reading
12 External links

Biography

**Early life and education**

Albert Einstein was born in Ulm, in the Kingdom of Württemberg in the German Empire on 14 March 1879.[8] His father was Hermann E based on direct current.[9]

The Einsteins were non-observant Jews. Albert attended a Catholic elementary school from the age of five for three years. Later, at the is disputed by the Albert Einstein Archives, and he excelled at the first school that he attended.[11]

His father once showed him a pocket compass; Einstein realized that there must be something causing the needle to move, despite th from Poland, was introduced to the Einstein family by his brother, and during weekly visits over the next five years he gave the boy pop

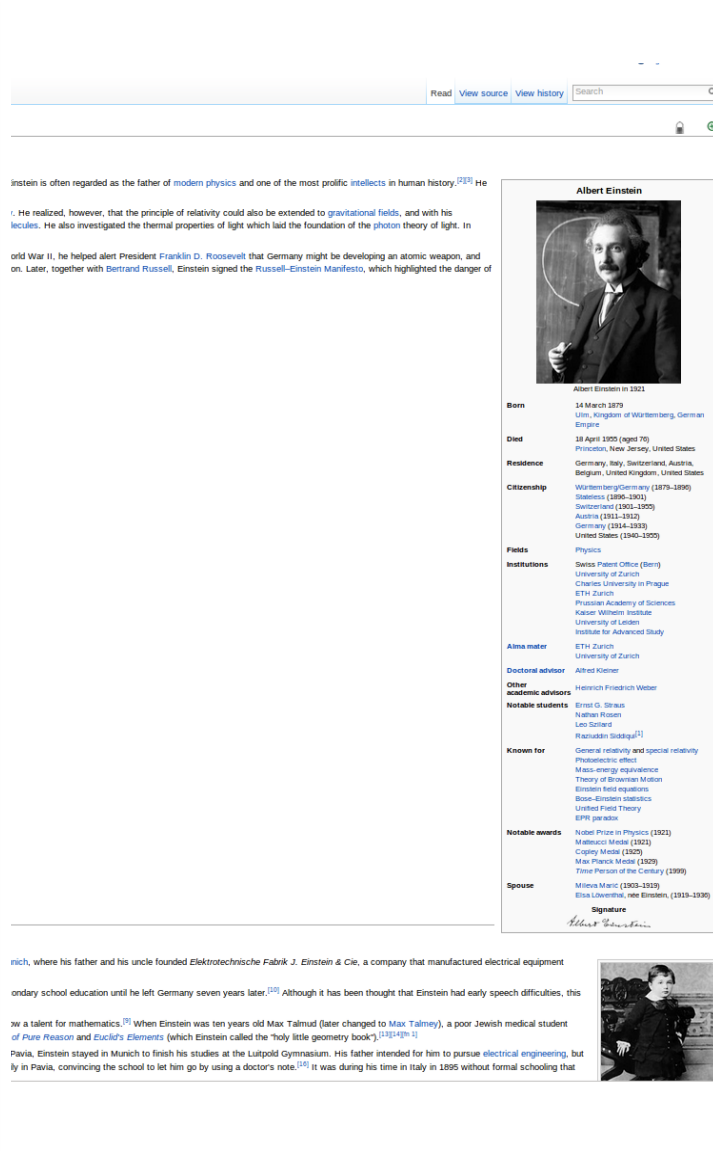
In 1894, his father's company failed: direct current (DC) lost the War of Currents to alternating current (AC). In search of business, the Einstein clashed with authorities and resented the school's regimen and teaching method. He later wrote that the spirit of learning and c

**Albert Einstein**



Albert Einstein in 1921

| | |
|---|---|
| **Born** | 14 March 1879<br>Ulm, Kingdom of Württemberg, German Empire |
| **Died** | 18 April 1955 (aged 76)<br>Princeton, New Jersey, United States |
| **Residence** | Germany, Italy, Switzerland, Austria, Belgium, United Kingdom, United States |
| **Citizenship** | Württemberg/Germany (1879–1896)<br>Stateless (1896–1901)<br>Switzerland (1901–1955)<br>Austria (1911–1912)<br>Germany (1914–1933)<br>United States (1940–1955) |
| **Fields** | Physics |
| **Institutions** | Swiss Patent Office (Bern)<br>University of Zurich<br>Charles University in Prague<br>ETH Zurich<br>Prussian Academy of Sciences<br>Kaiser Wilhelm Institute<br>University of Leiden<br>Institute for Advanced Study |
| **Alma mater** | ETH Zurich<br>University of Zurich |
| **Doctoral advisor** | Alfred Kleiner |
| **Other academic advisors** | Heinrich Friedrich Weber |
| **Notable students** | Ernst G. Straus<br>Nathan Rosen<br>Leo Szilard<br>Raziuddin Siddiqui[1] |
| **Known for** | General relativity and special relativity<br>Photoelectric effect<br>Mass-energy equivalence<br>Theory of Brownian Motion<br>Einstein field equations<br>Bose–Einstein statistics |



Read View source View history Search

instein is often regarded as the father of modern physics and one of the most prolific intellects in human history.[2][3] He

r. He realized, however, that the principle of relativity could also be extended to gravitational fields, and with his lecules. He also investigated the thermal properties of light which laid the foundation of the photon theory of light. In

orld War II, he helped alert President Franklin D. Roosevelt that Germany might be developing an atomic weapon, and on. Later, together with Bertrand Russell, Einstein signed the Russell–Einstein Manifesto, which highlighted the danger of

nich, where his father and his uncle founded *Elektrotechnische Fabrik J. Einstein & Cie*, a company that manufactured electrical equipment

ondary school education until he left Germany seven years later.[10] Although it has been thought that Einstein had early speech difficulties, this

ow a talent for mathematics.[5] When Einstein was ten years old Max Talmud (later changed to Max Talmey), a poor Jewish medical student *of Pure Reason* and *Euclid's Elements* (which Einstein called the "holy little geometry book").[13][14][n 1]

Pavia, Einstein stayed in Munich to finish his studies at the Luitpold Gymnasium. His father intended for him to pursue electrical engineering, but ly in Pavia, convincing the school to let him go by using a doctor's note.[14] It was during his time in Italy in 1895 without formal schooling that

Albert Einstein

Albert Einstein in 1921

Born 14 March 1879 Ulm, Kingdom of Württemberg, German Empire

Died 18 April 1955 (aged 76) Princeton, New Jersey, United States

Residence Germany, Italy, Switzerland, Austria, Belgium, United Kingdom, United States

Citizenship Württemberg/Germany (1879–1896) Stateless (1896–1901) Switzerland (1901–1955) Austria (1911–1912) Germany (1914–1933) United States (1940–1955)

Fields Physics

Institutions Swiss Patent Office (Bern) University of Zurich Charles University in Prague ETH Zurich Prussian Academy of Sciences Kaiser Wilhelm Institute University of Leiden Institute for Advanced Study

Alma mater ETH Zurich University of Zurich

Doctoral advisor Alfred Kleiner

Other academic advisors Heinrich Friedrich Weber

Notable students Ernst G. Straus Nathan Rosen Leo Szilard Raziuddin Siddiqui[1]

Known for General relativity and special relativity Photoelectric effect Mass-energy equivalence Theory of Brownian Motion Einstein field equations Bose–Einstein statistics Unified Field Theory EPR paradox

Notable awards Nobel Prize in Physics (1921) Matteucci Medal (1921) Copley Medal (1925) Max Planck Medal (1929) *Time* Person of the Century (1999)

Spouse Mileva Marić (1903–1919) Elsa Löwenthal, née Einstein, (1919–1936)

Signature

# Motivation



- Creating new structured data sources (knowledge bases)
  - DBPedia
  - Freebase
  - Yago
- Answering complex questions using multiple sources

Which soccer player married a Spice Girls star?

```
("?x"    is-a      "soccer player")
("?x"    married   "?y")
("?y"    member    "Spice Girls")
```

# Outline

# Relation Representation



- Representing data as triples

  (Argument1 RelationType Argument2)
  (Subject Predicate Object)

- Resource Description Framework (RDF)

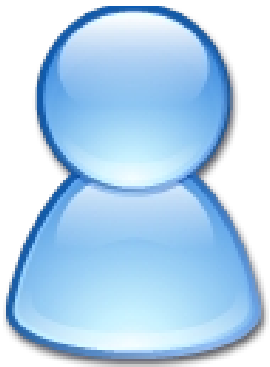
# Relation Types



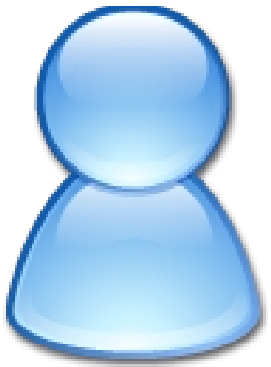
- Having various relation types based on the type of arguments
  - PER-PER
  - ORG-LOC
  - PER-ORG
  - PER-LOC
  - PER-DATE
  - Drugs-Disease

# Relation Types



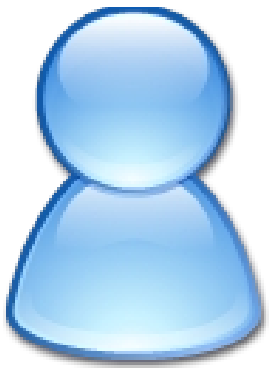


Spouse
Parent
Child
Friendship
Colleague
..

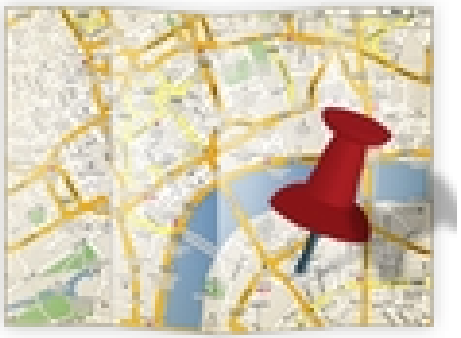


# Relation Types



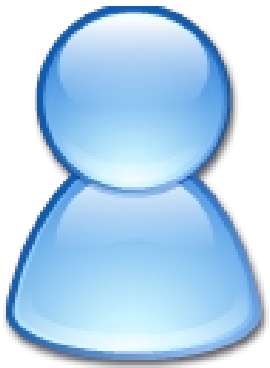


Place of birth
Place of death
Buried in
..



# Relation Types



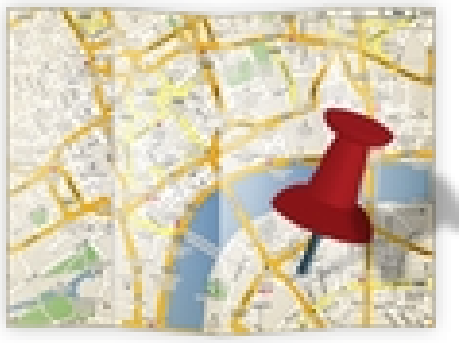


Founder
Co-founder
Owner
Employee
Student/Alum
Professor
..

# Relation Types



Located
Near
Founded-location
Headquarter
..



# Approaches



- Manually created patterns
- Supervised machine learning
- Semi-supervised learning

# Outline

# Pattern Extraction

- What are the potential words to express a relation type?
  (PER Member ORG)
  ("?x" Member "?y")

x is a member of y.
x is an employee of y.
x works at y.
x is a staff of y.

x is (a|an) (member|employee|staff) of y.
x (works) at y.

# Pattern Extraction



- Advantages
  - Having high precision results
- Disadvantages
  - Having low recall
  - Finding all possible patterns is labor intensive
  - Covering all relations is very difficult

# Outline

# Supervised Classification



- Training data
  - Defining a fix set of relation types
  - Choosing the corresponding named entities
  - Selecting a set of text as training data
  - Recognizing the named entities in the text
  - Labeling the relations between named entities manually

# Task

- Input
  - A pair of entities
  - A context that this pair appears in
  - Possible relation types
- Output
  - Type of relation between two entities, if there exist any

Thomas Edison

Thomas Edison died on October 18, 1931,
in New Jersey due to complications of diabetes.

Place of birth
Place of death ✔
Buried in
..

New Jersey

# Feature Selection



Thomas Edison died on October 18, 1931, in New Jersey due to complications of diabetes.

- The target entities
  - $T1$: Thomas Edison
  - $T2$: New Jersey
- Surrounding words of target entities
  - $T1_{+1}$: died
  - $T2_{-1}$: in
  - $T2_{+1}$: due
- All words between the target entities (bag-of-word)
  - died
  - on
  - October
  - 18
  - ,
  - 1931
  - ,
  - in

# Feature Selection

Thomas Edison died on October 18, 1931, in New Jersey due to complications of diabetes.

- Other Features
  - The named entity label of the target words
    - $NE(T1)$: PER
    - $NE(T2)$: LOC
- The syntactic structure of the words between target words

# Classification Algorithm



- Applying any kinds of classifiers
  - *K* Nearest Neighbor
  - Support Vector Machines
  - Naïve Bayes
  - Maximum Entropy
  - Logistic Regression
  - ...

# Supervised Classification



- Advantages
  - Very good performance if
    - Having enough training data
    - Having test data similar to training data

- Disadvantages
  - Manual labeling of training data is labor expensive
  - Difficult to get good results for other genres

# Outline

# Semi-supervised Learning



- Having no large training data
- Producing a small training data (seed data)
  - A set of triples
- Using the seed data to find further entity pairs with the same relation

Bootstrapping

# Bootstrapping



- Using the collected seed data
- Finding sentences which contain at least one entity pairs
- Extracting the common contexts of the pair
- Creating patterns from the extracted context
- Using the pattern to grep more pairs and add them to seed data

# Bootstrapping



- Using the collected seed data

( Thomas Edison  Spouse  Mina Mille)

# Bootstrapping



- Using the collected seed data
- Finding sentences which contain at least one entity pairs

Thomas Edison married Mina Mille.
Edison married a young woman named Mina Mille.
In 1871, Thomas Edison married Mina Mille.
Thomas Edison marries Mina Mille on December 25.

# Bootstrapping



- Using the collected seed data
- Finding sentences which contain at least one entity pairs
- Extracting the common contexts of the pair
- Creating patterns from the extracted context

Thomas Edison married Mina Mille.
Edison married a young woman named Mina Mille.
In 1871, Thomas Edison married Mina Mille.
Thomas Edison marries Mina Mille on December 25.

# Bootstrapping



- Using the collected seed data
- Finding sentences which contain at least one entity pairs
- Extracting the common contexts of the pair
- Creating patterns from the extracted context
- Using the pattern to grep more pairs and add them to seed data

( Albert Eistein  Spouse  “?” )

Einstein marries his cousin Elsa Löwenthal on June 2.
Einstein married Elsa Löwenthal in Berlin.
Einstein married Elsa Löwenthal on 2 June 1919.
After their divorce in 1919, Einstein married Elsa Löwenthal in the same year.
Albert Einstein was married to Elsa Löwenthal for 17 years.
Einstein marries Elsa Löwenthal.
In the same year Albert Einstein married Elsa Löwenthal.

⇒ ( Albert Eistein  Spouse  Elsa Löwenthal )

# Bootstrapping



- Using the collected seed data

( Thomas Edison Spouse Mina Mille)

( Albert Eistein Spouse Elsa Löwenthal )

# Bootstrapping



- Using the collected seed data
- Finding sentences which contain at least one entity pairs
- Extracting the common contexts of the pair
- Creating patterns from the extracted context

Albert Einstein's wife, Elsa Löwenthal, was his first cousin.
Elsa Löwenthal was the wife of Albert Einstein.
Einstein's wife was named Elsa Löwenthal.

# Further Reading



- Speech and Language Processing
  - Chapters 22.2