(Open) Information Extraction: Overview and Recent Advances

Claudio Delli Bovi

Dipartimento di Informatica

Sapienza Università di Roma

Where do we go now?

You’ve seen many powerful methods for NLP:

- Language models, Part-of-Speech tagging, Syntactic Parsing, …
- Word Sense Disambiguation, Machine Translation, Semantic Similarity, ...

Where do we go now?

You've seen many powerful methods for NLP:

- Language models, Part-of-Speech tagging, Syntactic Parsing, …
- Word Sense Disambiguation, Machine Translation, Semantic Similarity, ...

The goal: Machine Reading

- Autonomous **understanding** of **unstructured** text on a **large scale**

Where do we go now?

You've seen many powerful methods for NLP:

- Language models, Part-of-Speech tagging, Syntactic Parsing, …
- Word Sense Disambiguation, Machine Translation, Semantic Similarity, ...

The goal: Machine Reading

- Autonomous **understanding** of **unstructured** text on a **large scale**

"*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 across the entire English-speaking web, and placing those facts in a structured knowledge base.*"
(T. Mitchell. Reading the Web: A Breakthrough Goal for AI. AI Magazine, 2005)

Where do we go now?

Machine Reading is an **ill-posed** problem:

Human Reading

✓ high precision
✓ broad scope
✓ high comprehension
✓ background knowledge
× sentence-by-sentence
× (usually) single language
× slow

Machine Reading

× noisy
× limited scope
× minimal reasoning
× bottom up
✓ corpus-wide statistics
✓ (possibly) multilingual
✓ very fast!

O. Etzioni, M. Banko, M.J. Cafarella. Machine Reading. AAAI, 2007.

Information Extraction

“A process of getting **structured** data from **unstructured** information in the text”
(Jurafsky and Martin, 2009)

“Identification of instances of a particular class of **relationships** in a natural language text, and the extraction of relevant **arguments** for that relationships”
(Grishman, 1997)

Information Extraction

Traditional Information Extraction pipeline:

Citing high fuel prices, United Airlines said Friday it has increased fares by $6 per round trip on flights to some cities also served by lower-cost carriers. American Airlines, a unit of AMR Corp., immediately matched the move, spokesman Tim Wagner said. United, a unit of UAL Corp., said the increase took effect Thursday and applies to most routes where it competes against discount carriers, such as Chicago to Dallas and Denver to San Francisco.

(Jurafsky and Martin, 2009)

Information Extraction

Traditional Information Extraction pipeline:

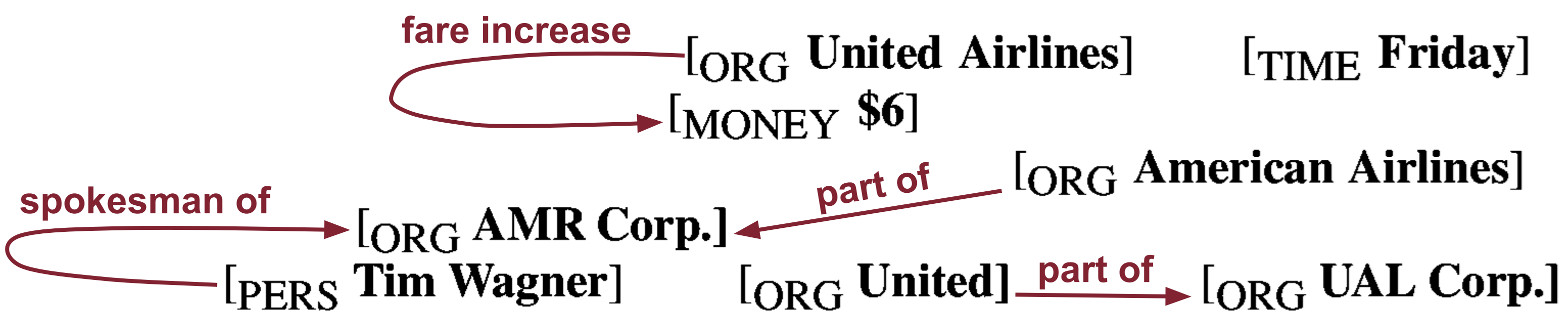
Citing high fuel prices, [$_{\text{ORG}}$ **United Airlines**] said [$_{\text{TIME}}$ **Friday**] it has increased fares by [$_{\text{MONEY}}$ **$6**] per round trip on flights to some cities also served by lower-cost carriers. [$_{\text{ORG}}$ **American Airlines**], a unit of [$_{\text{ORG}}$ **AMR Corp.**], immediately matched the move, spokesman [$_{\text{PERS}}$ **Tim Wagner**] said. [$_{\text{ORG}}$ **United**], a unit of [$_{\text{ORG}}$ **UAL Corp.**], said the increase took effect [$_{\text{TIME}}$ **Thursday**] and applies to most routes where it competes against discount carriers, such as [$_{\text{LOC}}$ **Chicago**] to [$_{\text{LOC}}$ **Dallas**] and [$_{\text{LOC}}$ **Denver**] to [$_{\text{LOC}}$ **San Francisco**].

(Jurafsky and Martin, 2009)

1. **Named Entity Recognition** (homework 3, anyone?)

Information Extraction

Traditional Information Extraction pipeline:



`( American Airlines , part of , AMR Corp. )`

`( Tim Wagner , spokesman of , AMR Corp. )`

`( United , part of , UAL Corp. )`

1. **Named Entity Recognition** (homework 3, anyone?)
2. **Relation Extraction**

Information Extraction

Yes, but… which (and how many) relations?

- Restricted set of semantic relations **handcrafted** by humans (either general or domain-specific)

Information Extraction

Yes, but… which (and how many) relations?

- Restricted set of semantic relations **handcrafted** by humans (either general or domain-specific)

- Wikipedia **infoboxes**!

WIKIPEDIA
The Free Encyclopedia

(Sapienza , Rector , Eugenio Gaudio)

(Sapienza , Location , Rome)

Sapienza University of Rome
Sapienza – Università di Roma

Latin: *Studium Urbis*

Motto	*Il futuro è passato qui*
Motto in English	*The future has passed here*
Established	1303
Type	Public
Rector	Dr. Eugenio Gaudio
Administrative staff	8,000
Students	112,564[1]
Location	Rome, Italy

Information Extraction

Yes, but… which (and how many) relations?

- Restricted set of semantic relations **handcrafted** by humans (either general or domain-specific)

- Wikipedia **infoboxes**!

DBpedia

http://wiki.dbpedia.org

Crowdsourced ontology derived (mainly) from Wikipedia infoboxes and containing over **2 billion** RDF triples.

Freebase

http://www.freebase.com

`people/person/nationality`

`location/location/contains`

`people/person/place-of-birth`

…

Information Extraction

Yes, but… which (and how many) relations?

- Restricted set of semantic relations **handcrafted** by humans (either general or domain-specific)

- Wikipedia **infoboxes**

- Ontological relations from thesauri like **WordNet**

  is a, instance of (hypernymy)

  part of (meronymy)

  ...

  entailment

  pertanymy

Information Extraction

Plenty of approaches:

Hand-written patterns and rules

Supervised learning algorithms

Semi-supervised learning algorithms

Weak and distant supervision

Unsupervised algorithms
(Open Information Extraction)

Information Extraction

Hand-written patterns and rules (Hearst, 1992)

$$\text{NP } \{, \text{NP}\}^* \{,\} (\text{and}|\text{or}) \text{ other } \text{NP}_H$$
$$\text{NP}_H \text{ such as } \{\text{NP},\}^* \{(\text{or}|\text{and})\} \text{ NP}$$
$$\text{such } \text{NP}_H \text{ as } \{\text{NP},\}^* \{(\text{or}|\text{and})\} \text{ NP}$$
$$\text{NP}_H \{,\} \text{ including } \{\text{NP},\}^* \{(\text{or}|\text{and})\} \text{ NP}$$
$$\text{NP}_H \{,\} \text{ especially } \{\text{NP}\}^* \{(\text{or}|\text{and})\} \text{ NP}$$

(Jurafsky and Martin, 2009)

Information Extraction

Supervised Learning (Zhao and Grishman, 2005; Bunescu and Mooney, 2006)

Information Extraction

Supervised Learning (Zhao and Grishman, 2005; Bunescu and Mooney, 2006)

- Start from a **fixed set** of relations and entities

Information Extraction

Supervised Learning (Zhao and Grishman, 2005; Bunescu and Mooney, 2006)

- Start from a **fixed set** of relations and entities
- Use these to annotate a large enough **training corpus**

  Positive examples: annotated triples ⟨entity, relation, entity⟩

  Negative examples: generated from non-annotated within-sentence entity pairs

Information Extraction

Supervised Learning (Zhao and Grishman, 2005; Bunescu and Mooney, 2006)

- Start from a **fixed set** of relations and entities
- Use these to annotate a large enough **training corpus**
  - **Positive examples: annotated triples ⟨entity, relation, entity⟩**
  - **Negative examples: generated from non-annotated within-sentence entity pairs**
- Train a **classifier** to annotate unseen text
  - **Word features (bag-of-words, headwords, bigrams…)**
  - **NE features (entity types and their concatenation)**
  - **Syntactic features (constituents, dependency paths, ...)**

Information Extraction

Semi-supervised/Weakly Supervised Learning (Kozareva and Hovy, 2010)

Information Extraction

Semi-supervised/Weakly Supervised Learning (Kozareva and Hovy, 2010)

“ [...] *Ryanair has a hub at Charleroi.”*

`(Ryanair, hub, Charleroi)`

Start from very few high-precision **seed patterns** (or **seed triples**)

Information Extraction

Semi-supervised/Weakly Supervised Learning (Kozareva and Hovy, 2010)

“ [...] Ryanair has a hub at Charleroi.”

`(Ryanair, hub, Charleroi)`

WIKIPEDIA
The Free Encyclopedia

“Budget airline **Ryanair**, which uses **Charleroi** as a **hub**, scrapped all weekend flights out of the airport”

“All flights in and out of **Ryanair**’s Belgian **hub** at **Charleroi** airport were grounded on Friday”

“A spokesman at **Charleroi**, a main **hub** for **Ryanair**, estimated that 8000 passengers had already been affected”

Bootstrapping:

1. Find sentences that contain both entities

Information Extraction

Semi-supervised/Weakly Supervised Learning (Kozareva and Hovy, 2010)

“ [...] Ryanair has a hub at Charleroi.”

(Ryanair, hub, Charleroi)

WIKIPEDIA
The Free Encyclopedia

“Budget airline **Ryanair**, which uses **Charleroi** as a **hub**, scrapped all weekend flights out of the airport”

“All flights in and out of **Ryanair**’s Belgian **hub** at **Charleroi** airport were grounded on Friday”

“A spokesman at **Charleroi**, a main **hub** for **Ryanair**, estimated that 8000 passengers had already been affected”

Bootstrapping:

1. Find sentences that contain both entities
2. Generalize to new patterns

```
/[ORG], which uses [LOC] as a hub/
/[ORG]'s hub at [LOC]/
/[LOC] a main hub for [ORG]/
```

Information Extraction

Self-supervised Learning (Etzioni et al., 2005; Weld et al., 2008)

Idea: use the new patterns to search for additional triples and build a self-labeled training dataset

Information Extraction

Self-supervised Learning (Etzioni et al., 2005; Weld et al., 2008)

Idea: use the new patterns to search for additional triples and build a self-labeled training dataset

Issues: error propagation, semantic drift (erroneous patterns leads to the introduction of erroneous tuples, which, in turn...)

`/[ORG] has a hub at [LOC]/`

“ [...] *Sydney has a ferry hub at Circular Quay”*

`(Sydney, hub, Circular Quay)`

?

Information Extraction

NELL - Never Ending Language Learning (Carlson et al., 2010)

Web-scale self-supervised learning system, running at CMU continuously 24 hours per day

Read the Web

Research Project at Carnegie Mellon University

Information Extraction

NELL - Never Ending Language Learning (Carlson et al., 2010)

Web-scale self-supervised learning system, running at CMU continuously 24 hours per day

Requires:

- An initial **ontology** with categories (`person,sportsTeam, fruit, ...`) and relations (`playsInstrument, playsOnTeam, ...`)

Information Extraction

NELL - Never Ending Language Learning (Carlson et al., 2010)

Web-scale self-supervised learning system, running at CMU continuously 24 hours per day

Requires:

- An initial **ontology** with categories (**`person,sportsTeam, fruit, ...`**) and relations (**`playsInstrument, playsOnTeam, ...`**)
- 10-15 **seed examples** of each category and relation

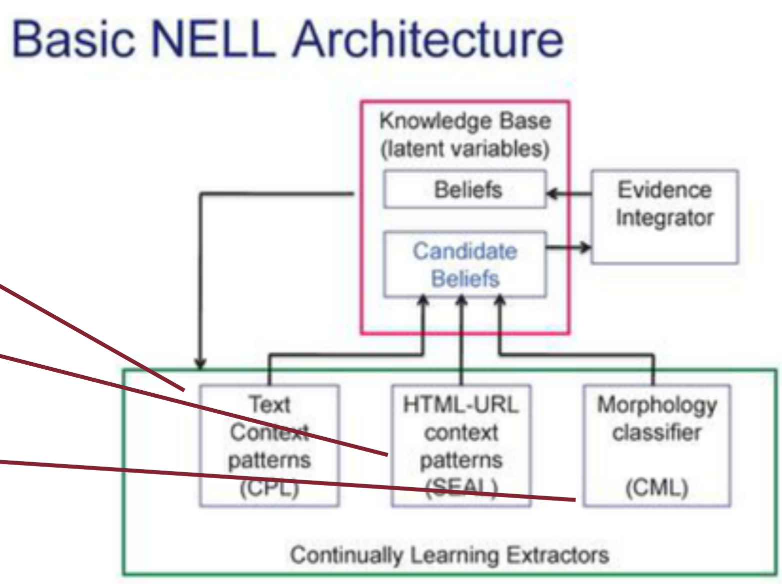
Information Extraction

NELL - Never Ending Language Learning (Carlson et al., 2010)

Uses a variety of methods to extract **beliefs** from the web

Hundreds of different extraction modules simultaneously trained:

- **Co-occurrence based pattern learners**
- **HTML lists and tables miners**
- **“Traditional” classifiers on various features**



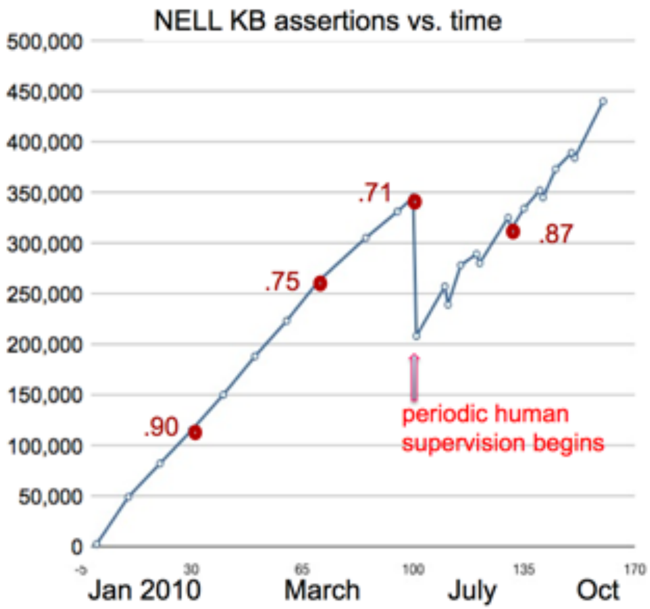
Information Extraction

NELL - Never Ending Language Learning (Carlson et al., 2010)

Uses a variety of methods to extract **beliefs** from the web

Hundreds of different extraction modules simultaneously trained...

..and a bit of human supervision once in a while!



Information Extraction

NELL - Never Ending Language Learning (Carlson et al., 2010)

Recently-Learned Facts twitter Refresh

instance	iteration	date learned	confidence
valve_of_navicular_fossa is a nerve	925	17-may-2015	99.4
agartala is a visualizable scene	922	05-may-2015	98.2
kings_paget_hotel_west_drayton is a hotel	922	05-may-2015	100.0
workers_in_struggle_collectives is a trade union	926	20-may-2015	91.4
kids_clothing_sets is a household item	923	08-may-2015	97.9
jungle_cat_world is an aquarium in the city orono	927	26-may-2015	100.0
flowers is an agricultural product produced in croatia	925	17-may-2015	100.0
reuters is headquartered in the city new_york	927	26-may-2015	100.0
national_hockey_league participated in the event series	922	05-may-2015	98.4
winter is an organization dissolved at the date may	923	08-may-2015	99.9

http://rtw.ml.cmu.edu/rtw/

Information Extraction

Distantly Supervised Learning (Mintz et al., 2009; Riedel et al., 2010, Hoffmann et al., 2011)

Instead of just a handful of seeds, use a **large knowledge base** (possibly human-contributed) to acquire many reliable training examples

Information Extraction

Distantly Supervised Learning (Mintz et al., 2009; Riedel et al., 2010, Hoffmann et al., 2011)

Instead of just a handful of seeds, use a **large knowledge base** (possibly human-contributed) to acquire many reliable training examples

For each entity pair, **identify all sentences** mentioning them in a massive unlabeled corpus

Information Extraction

Distantly Supervised Learning (Mintz et al., 2009; Riedel et al., 2010, Hoffmann et al., 2011)

Instead of just a handful of seeds, use a **large knowledge base** (possibly human-contributed) to acquire many reliable training examples

For each entity pair, **identify all sentences** mentioning them in a massive unlabeled corpus

Lots of noisy pattern features, then combined in a **supervised classifier**

Information Extraction

Distantly Supervised Learning (Mintz et al., 2009; Riedel et al., 2010, Hoffmann et al., 2011)

Distant supervision assumption:

If two entities participate in a relation, **all sentences** that mention these two entities express that relation

Information Extraction

Distantly Supervised Learning (Mintz et al., 2009; Riedel et al., 2010, Hoffmann et al., 2011)

Distant supervision assumption:

If two entities participate in a relation, **all sentences** that mention these two entities express that relation

`(Brad Pitt, married with, Angelina Jolie)`

*" **Brad Pitt** will be starring with **Angelina Jolie** in 'World War X' "*

*"**Angelina Jolie** joins **Brad Pitt** for first public appearance [...]"*

*"**Angelina Jolie** expects another baby with **Brad Pitt**"*

Information Extraction

Distantly Supervised Learning (Mintz et al., 2009; Riedel et al., 2010, Hoffmann et al., 2011)

Relax the assumption:

- If two entities participate in a relation, ~~all sentences~~ **at least one sentence** that mention these two entities express that relation
  (*expressed-at-least-once* assumption)

Multi-instance

Information Extraction

Distantly Supervised Learning (Mintz et al., 2009; Riedel et al., 2010, Hoffmann et al., 2011)

Relax the assumption:

- If two entities participate in a relation, ~~all sentences~~ **at least one sentence** that mention these two entities express that relation
- Allow for some relations to **overlap**:

`(Steve Jobs, founded, Apple)` `(Steve Jobs, CEO of, Apple)`

Multi-instance Multi-label Learning (Surdeanu et al., 2012)

Open Information Extraction

Unsupervised Learning (Banko and Etzioni, 2008; Wu and Weld, 2010)

A way more radical approach:

no predefined set of relations

no human intervention

no training data

only a large unlabeled corpus (like the Web) as input

Open Information Extraction

"Open Information Extraction (OIE) [...] a novel extraction paradigm that facilitates **domain-independent** discovery of relations extracted from text and readily **scales to the diversity and size of the Web** corpus. The sole input to an OIE system is a corpus, and its output is a set of extracted relations. An OIE system makes a single pass over its corpus guaranteeing scalability with the size of the corpus."

(Banko et al., 2007)

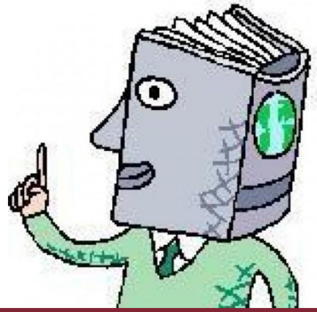
Web-scale unlabeled corpus

OIE system

Unstructured set of relation instances (triples)

Open Information Extraction

ReVerb (Fader et al., 2011)

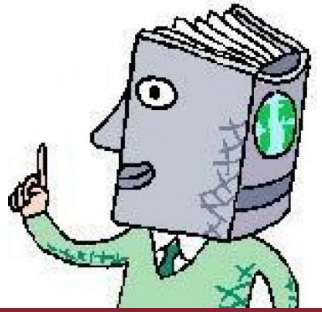


Open Information Extraction

ReVerb (Fader et al., 2011)

Given a sentence *s*:

- POS tagging and chunking over *s;*

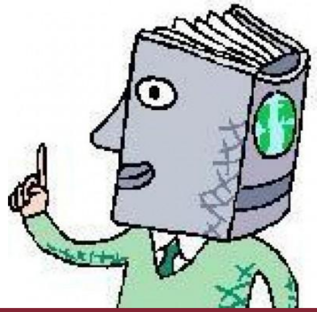


Open Information Extraction

ReVerb (Fader et al., 2011)

Given a sentence *s*:

- POS tagging and chunking over *s;*
- For each **verb** in *s*, find the longest sequence of words *w* that starts with that verb and satisfies **syntactic** and **lexical constraints**;

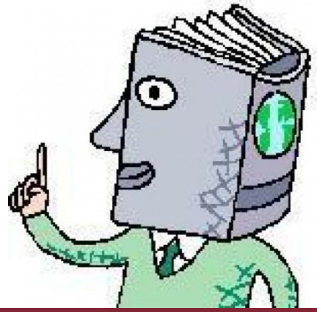


Open Information Extraction

ReVerb (Fader et al., 2011)

Given a sentence *s*:

- POS tagging and chunking over *s;*
- For each **verb** in *s*, find the longest sequence of words *w* that starts with that verb and satisfies **syntactic** and **lexical constraints**;
- For each phrase *w*, find the **nearest NP** to the left and to the right;

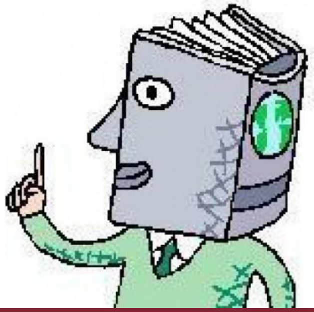


Open Information Extraction

ReVerb (Fader et al., 2011)

Given a sentence *s*:

- POS tagging and chunking over *s;*
- For each **verb** in *s*, find the longest sequence of words *w* that starts with that verb and satisfies **syntactic** and **lexical constraints**;
- For each phrase *w*, find the **nearest NP** to the left and to the right;
- Assign a confidence *c* to the relation *r* = *(x, w, y)* using a confidence classifier.



Open Information Extraction

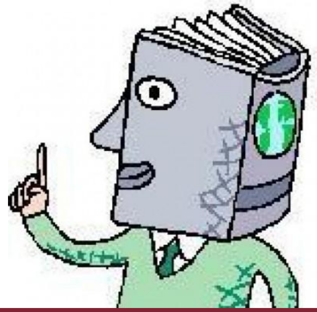
ReVerb (Fader et al., 2011)

s: *United has a hub in Chicago, which is the headquarters of United Continental Holdings.*

⇩

r_1: (United, has a hub in, Chicago)

r_2: (Chicago, is the headquarters of, United Continental Holdings)



Open Information Extraction

ReVerb (Fader et al., 2011)

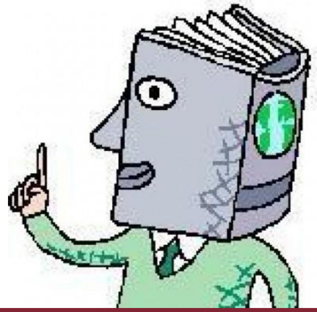
s: *United has a hub in Chicago, which is the headquarters of United Continental Holdings.*

⇩

r_1: **(United, has a hub in, Chicago)**

r_2: **(Chicago, is the headquarters of, United Continental Holdings)**

Almost **15 million extractions** (1.3 million distinct relations) from the **ClueWeb09** dataset!



Open Information Extraction

OIE is great, but...

Open Information Extraction

OIE is great, but...

Sparsity: many relation phrases actually express the same relationship (e.g. synonyms, paraphrases)

Open Information Extraction

OIE is great, but...

Sparsity: many relation phrases actually express the same relationship (e.g. synonyms, paraphrases)

Arguments (and relation phrases) are **ambiguous**

Open Information Extraction

OIE is great, but...

Sparsity: many relation phrases actually express the same relationship (e.g. synonyms, paraphrases)

Arguments (and relation phrases) are **ambiguous**

We need semantics!

Open Information Extraction

PATTY (Nakashole et al., 2012)

Open Information Extraction

PATTY (Nakashole et al., 2012)

From patterns to **pattern synsets** (clusters of relation phrases that express the same relation):

```
{ settled in , live in , moved to , stayed in ,
in area of , ...}
```

Open Information Extraction

PATTY (Nakashole et al., 2012)

From patterns to **pattern synsets** (clusters of relation phrases that express the same relation)

Each pattern synset (= relation) has **semantic types**:

`film/actor`	`already played with`	`film/actor`
`music/artist`	`already played with`	`music/composer`

`...`

Open Information Extraction

PATTY (Nakashole et al., 2012)

From patterns to **pattern synsets** (clusters of relation phrases that express the same relation)

Each pattern synset (= relation) has **semantic types**

Patterns are hierarchically organized in a **taxonomy**:

`{ is romantically involved with , is dating }` $\subset$ `{ knows , ... }`

Open Information Extraction

PATTY (Nakashole et al., 2012)

Generalized **syntactic-ontological-lexical** (**SOL**) patterns:

Amy Winehouse's soft voice in 'Rehab'

Elvis Presley's solid voice in his song 'All shook up'

Open Information Extraction

PATTY (Nakashole et al., 2012)

Generalized **syntactic-ontological-lexical** (**SOL**) patterns:

Amy Winehouse*'s soft voice in '****Rehab***'

Elvis Presley*'s solid voice in his song '****All shook up***'

select knowledge

Detect and link entities

Open Information Extraction

PATTY (Nakashole et al., 2012)

Generalized **syntactic-ontological-lexical** (**SOL**) patterns:

<Person> ’s [ADJ] voice * <Song>

Generalize SOL patterns

Amy Winehouse’*s soft voice in ‘****Rehab****’*

Elvis Presley’*s solid voice in his song ‘****All shook up****’*

Open Information Extraction

PATTY (Nakashole et al., 2012)

Generalized **syntactic-ontological-lexical** (**SOL**) patterns:

<Person> 's [ADJ] voice * <Song>

Amy Winehouse'*s soft voice in* '***Rehab***'

Elvis Presley'*s solid voice in his song* '***All shook up***'

(Amy Winehouse, Rehab)

(Elvis Presley, All Shook Up)

Extract relation instances (support set)

Open Information Extraction

PATTY (Nakashole et al., 2012)

Generalized **syntactic-ontological-lexical** (**SOL**) patterns:

`<Person> ’s [ ADJ ] voice * <Song>`

Lexical word features (**L**) harvested from an input corpora

Open Information Extraction

PATTY (Nakashole et al., 2012)

Generalized **syntactic-ontological-lexical** (**SOL**) patterns:

```
<Person> ’s [ ADJ ] voice * <Song>
```

Syntactic modifiers (**S**) are generalized using POS tags or wildcards

Open Information Extraction

PATTY (Nakashole et al., 2012)

Generalized **syntactic-ontological-lexical** (**SOL**) patterns:

<Person> ’s [ADJ] voice * <Song>

Ontological semantic types (**O**) from a knowledge base

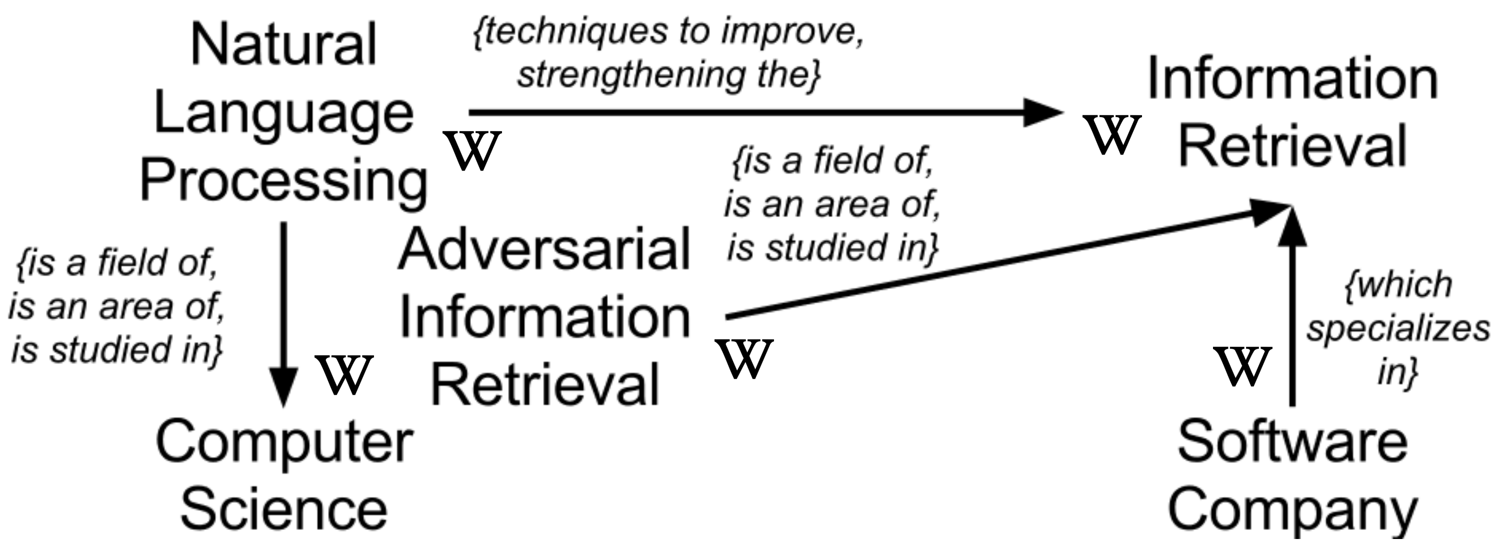
Open Information Extraction

WiSeNet (Moro and Navigli, 2012; 2013)

Open Information Extraction

WiSeNet (Moro and Navigli, 2012; 2013)

Wikipedia-based **Se**mantic **Net**work:



Open Information Extraction

WiSeNet (Moro and Navigli, 2012; 2013)

Filtering out bad relational phrases:

Wiesbaden State Library is funded by the State of Hesse and located in Wiesbaden.

Open Information Extraction

WiSeNet (Moro and Navigli, 2012; 2013)

Filtering out bad relational phrases:

Wiesbaden State Library *is funded by the* **State of Hesse** *and located in Wiesbaden.*

(**Wiesbaden State Library** , is funded by the , **State of Hesse**) ✓

Open Information Extraction

WiSeNet (Moro and Navigli, 2012; 2013)

Filtering out bad relational phrases:

Wiesbaden State Library *is funded by the State of Hesse and located in* **Wiesbaden***.*

(**Wiesbaden State Library** , is funded by the , **State of Hesse**) ✓

(**Wiesbaden State Library** , is funded by the State of Hesse and located in , **Wiesbaden**) ✓

Open Information Extraction

WiSeNet (Moro and Navigli, 2012; 2013)

Filtering out bad relational phrases:

Wiesbaden State Library is funded by the **State of Hesse** *and located in* **Wiesbaden***.*

(**Wiesbaden State Library** , is funded by the , **State of Hesse**) ✓

(**Wiesbaden State Library** , is funded by the State of Hesse and located in , **Wiesbaden**) ✓

(**State of Hesse** , and located in , **Wiesbaden**) ✗

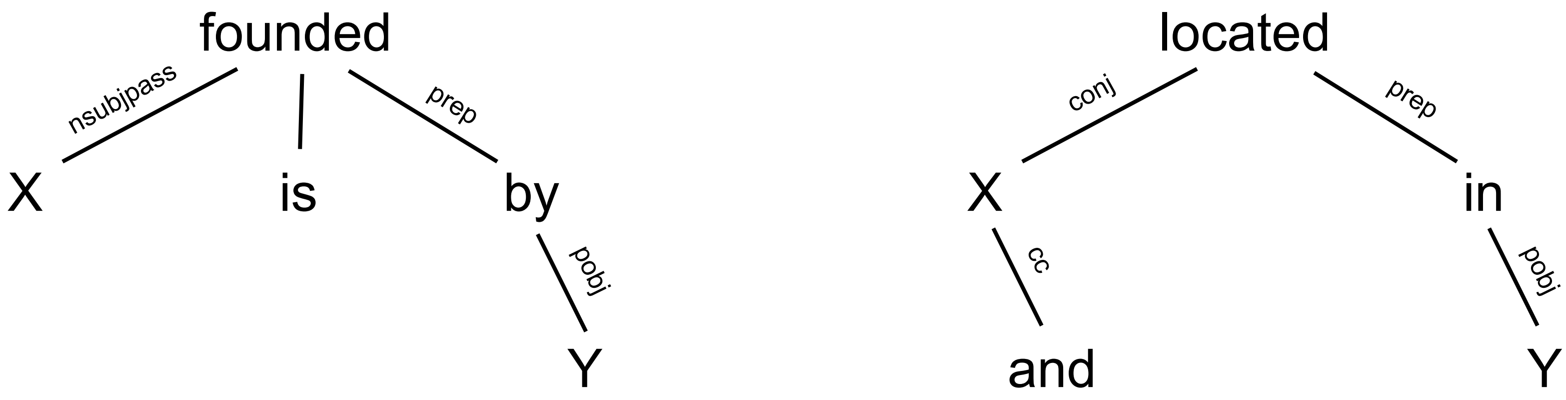
Open Information Extraction

WiSeNet (Moro and Navigli, 2012; 2013)

How? Use **syntactically-grounded** patterns:

X *is founded by* Y

X *and located in* Y

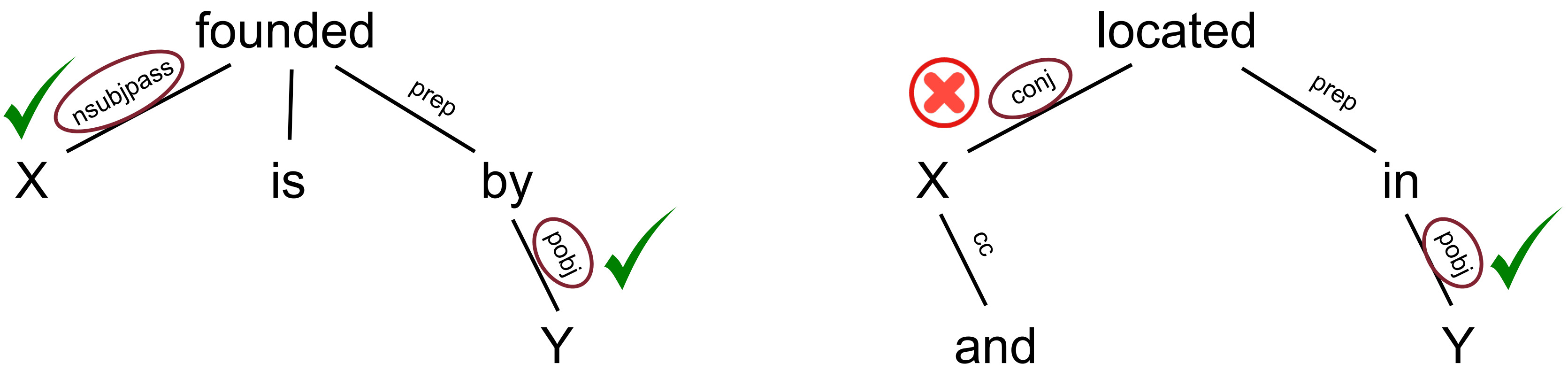


Open Information Extraction

WiSeNet (Moro and Navigli, 2012; 2013)

How? Use **syntactically-grounded** patterns:

X *is founded by* Y   X *and located in* Y



Open Information Extraction

WiSeNet (Moro and Navigli, 2012; 2013)

Use **soft clustering** techniques to build relation synsets with ambiguous patterns:

{ **is a part of** , is a territory of , … , is a province of }

{ **is a part of** , is a member of , … , is an element of }

{ made her acting debut in the , made his professional debut in the , … }

{ used to build, used to construct , … , used to manufacture }

Open Information Extraction

WiSeNet (Moro and Navigli, 2012; 2013)

Exploit Wikipedia categories to generate **semantic types**:

Domain	**Relation Synset**	**Range**
Arts	{is located in the small village of, . . ., is located in the small rural town of}	Places
Corporate groups	{is a member of an, . . ., were the members of the}	Corporate groups
Geography	{is a valley of, is a zone of, . . ., is a territory of}	Geography by place

Open Information Extraction

WiSeNet (Moro and Navigli, 2012; 2013)

Finally, use types to deal with **ambiguous relation instances**:

Ambiguous Relation Instance:		
(Natural language processing,	*is a field of,*	*Computer science)*

Open Information Extraction

WiSeNet (Moro and Navigli, 2012; 2013)

Finally, use types to deal with **ambiguous relation instances**:

Ambiguous Relation Instance:		
(Natural language processing,	*is a field of,*	*Computer science)*

Relation Synset Candidates:		
{Subfields by academic discipline, ..., Science}	{is a field of, is an area of, is studied in}	{Scientific Disciplines, ..., Science}
{Agriculture, ..., Horticulture and gardening	{is a field of, is cultivated with, where grows}	{Fruit, ..., Cultivars}
{Cities, ..., Villages}	{is a field of, was the site of, is the battlefield, }	{Battles, ..., Wars}

Open Information Extraction

WiSeNet (Moro and Navigli, 2012; 2013)

Finally, use types to deal with **ambiguous relation instances**:

Ambiguous Relation Instance:		
(Natural language processing,	*is a field of,*	*Computer science)*

Relation Synset Candidates:		
{Subfields by academic discipline, ..., Science}	{is a field of, is an area of, is studied in}	{Scientific Disciplines, ..., Science}
{Agriculture, ..., Horticulture and gardening	{is a field of, is cultivated with, where grows}	{Fruit, ..., Cultivars}
{Cities, ..., Villages}	{is a field of, was the site of, is the battlefield, }	{Battles, ..., Wars}

Semantic Relation:		
(Natural language processing,	{is a field of, is an area of, is studied in},	Computer science)

What else?

What else?

Put together syntactic and semantic analysis and generate **semantically augmented patterns:**

$$\text{X } \textit{directed by} \text{ Y}$$

$$\text{X } \textit{known for} \text{ Y}$$

$$\text{X } \textit{is} \ \texttt{election district}^{1}_{bn} \ \textit{of} \text{ Y}$$

$$\text{X } \textit{is} \ \texttt{composer}^{1}_{bn} \ \textit{from} \text{ Y}$$

$$\text{X } \textit{is} \ \texttt{street}^{1}_{bn} \ \textit{named after} \text{ Y}$$

What else?

election district · voting precinct · Electoral district · constituency

One of several districts into which a city or town is divided for voting; each contains one polling place

More definitions

IS-A: precinct • administrative division PART-OF: Representative democracy

EXPLORE NETWORK

composer

Someone who composes music as a profession More definitions

IS-A: musician • player • profession

EXPLORE NETWORK

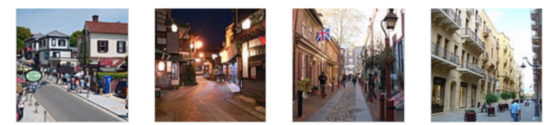
street

A thoroughfare (usually including sidewalks) that is lined with buildings

They walked the streets of the small town More examples

IS-A: thoroughfare • state highway • architectural structure

EXPLORE NETWORK



WORK IN PROGRESS

X directed by Y

X known for Y

X is $\boxed{\texttt{election district}^{1}_{bn}}$ *of Y*

X is $\boxed{\texttt{composer}^{1}_{bn}}$ *from Y*

X is $\boxed{\texttt{street}^{1}_{bn}}$ *named after Y*

So… happy (knowledge) harvesting!