Entity Linking - Part I

Pawan Goyal

CSE, IIT Kharagpur

Week 10, Lecture 1

Entity Linking: Introduction

- Entity Linking is the task of identifying all mentions in text of a specific entity from a database or ontology

Entity Linking: Introduction

- Entity Linking is the task of identifying all mentions in text of a specific entity from a database or ontology
- Researchers have used Wikipedia, YAGO, Freebase etc. as the databases

Entity Linking: Introduction

- Entity Linking is the task of identifying all mentions in text of a specific entity from a database or ontology
- Researchers have used Wikipedia, YAGO, Freebase etc. as the databases
- Typically broken down into two main phases
  - Candidate Selection (entity annotation)
  - Reference disambiguation or entity resolution

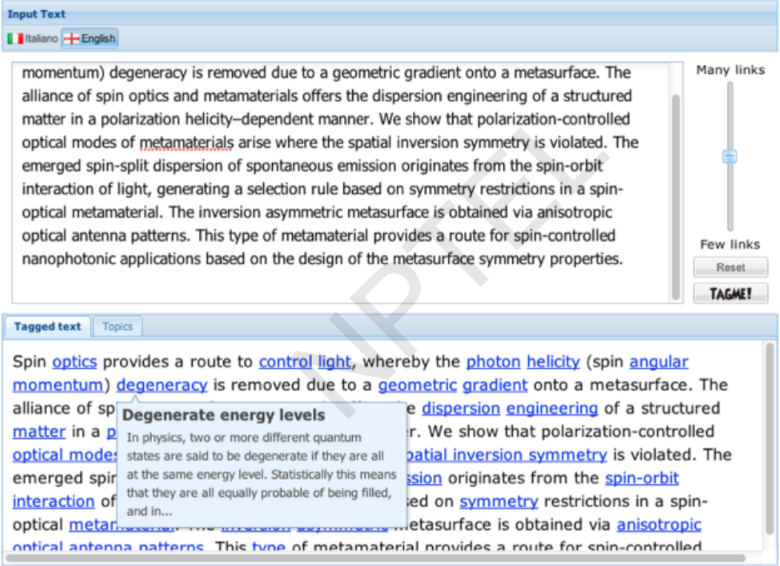
Entity Linking: Introduction

- Entity Linking is the task of identifying all mentions in text of a specific entity from a database or ontology
- Researchers have used Wikipedia, YAGO, Freebase etc. as the databases
- Typically broken down into two main phases
  - Candidate Selection (entity annotation)
  - Reference disambiguation or entity resolution
- Entitly linking needs to handle
  - Name variations (entities are referred to in many different ways)
  - Entity ambiguity (the same string can refer to more than one entity)

Entity Linking: Introduction

- We will take Wikipedia as the knowledge base to understand the task.
- With Wikipedia as the knowledge base, this task is commonly known as *Wikification*.

What is Entity Linking?



What is Entity Linking?

Iranian POW negotiator holds talks with Iraqi ministers

The head of **Iran**'s **prisoner of war** commission met with two **Iraqi** Cabinet ministers Saturday in a bid to glean information about thousands of Iranian POWs allegedly in Iraq, the official Iraqi News Agency reported.

Iraqi Foreign Minister **Mohammed Saeed al-Sahhaf** told Abdullah al-Najafi that the two states needed to ``speed up the closure of what remains from the POW and Missing-In-Action file,'' INA said.

The issue of POWs and missing persons remains a stumbling block to normalizing relations between the two neighbors.

Iraq has long maintained that it has released all Iranian prisoners captured in the **1980**-88 **Iran-Iraq War**. The countries accuse each other of hiding POWs and preventing visits by the **International Committee of the Red Cross** to prisoner camps.

The ICRC representative in **Baghdad**, Manuel Bessler, told The **Associated Press** that his organization has had difficulty visiting POWs on both sides on a regular basis.

In April, Iran released 5,584 … fied as civil law detainees in the largest exchange since **1990**.

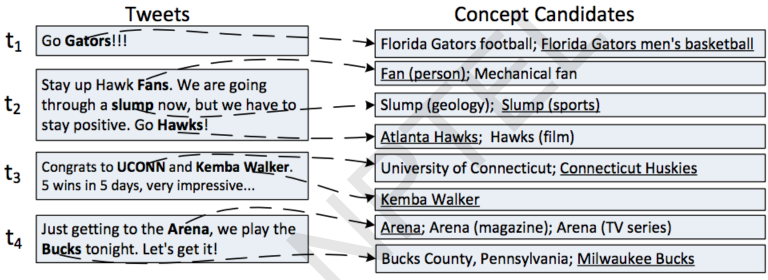
More than 1 million people w

Baghdad

Baghdad is the capital of Iraq and of Baghdad Governorate. With a metropolitan area estimated at a population of 7,000,000, it is the largest city in Iraq. It is the second-largest city in the Arab world (after Cairo) and the second-largest city in southwest Asia (after Tehran).

open in wikipedia

What is Entity Linking?



Entity Linking: Common Steps

Determine "linkable" phrases

mention detection - **MD**

Entity Linking: Common Steps

Determine "linkable" phrases

mention detection - **MD**

Rank/Select candidate entity links

link generation - **LG**

Entity Linking: Common Steps

Determine "linkable" phrases

mention detection - **MD**

Rank/Select candidate entity links

link generation - **LG**

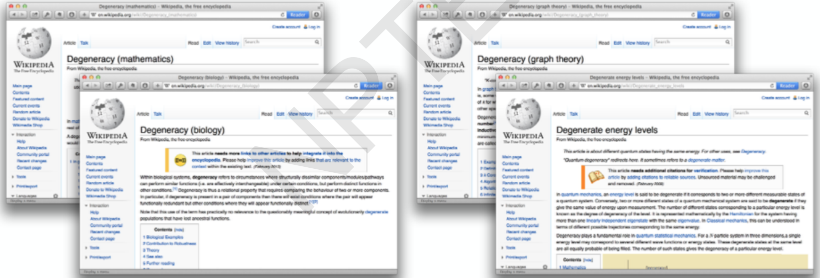
Use "context" to disambiguate/filter/improve

disambiguation - **DA**

Mention Detection (MD)

Link Generation (LG)

… degeneracy …



Disambiguation (DA)

Preliminaries: Wikipedia

- Basic element: article (proper)
- But also
  - redirect pages
  - disambiguation pages
  - category/template pages
  - admin pages
- Hyperlinks
  - use “unique identifiers” (URLs)
    - [[United States]] or [[United States|American]]
    - [[United States (TV series)]] or [[United States (TV series)|TV show]]

- Senses of an ambiguous phrase
- Short description
- (Possible) categorization
- Non-exhaustive

Some Statistics

WordNet

- 80k entity definitions
- 142k senses (entity - surface forms)

Some Statistics

WordNet

- 80k entity definitions
- 142k senses (entity - surface forms)

Wikipedia

- 4M entity definitions
- 24M senses

Wikipedia based methods

What can be a good measure for MD?

Wikipedia based methods

What can be a good measure for MD?

keyphraseness(w)

Number of Wikipedia articles that use it as an anchor, divided by the numer of articles that mention it at all.

Wikipedia based methods

What can be a good measure for MD?

keyphraseness(w)

Number of Wikipedia articles that use it as an anchor, divided by the numer of articles that mention it at all.

$$\frac{\mathrm{CF}(w_l)}{\mathrm{CF}(w)}$$

Collection frequency
term w as a link to another Wikipedia article

Collection frequency
term w

Wikipedia based methods

What can be a good measure for DA?

Wikipedia based methods

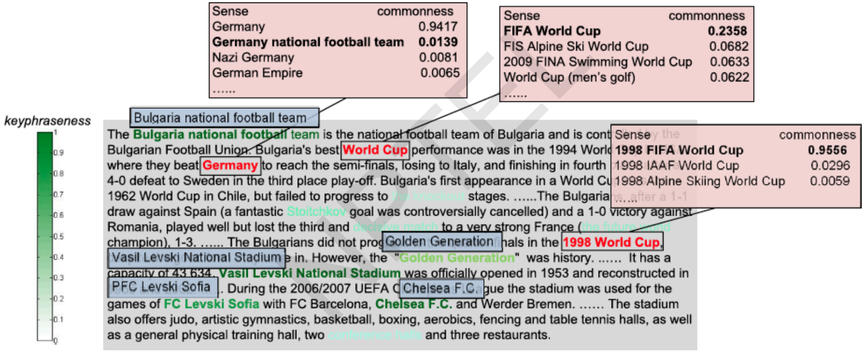
What can be a good measure for DA?

commonness(w, c)

The fraction of times, a particular sense is used as a destination in Wikipedia.

$$\frac{|L_{w,c}|}{\sum_{c'} \underbrace{|L_{w,c'}|}_{\substack{\textbf{Number of links} \\ \text{with target c' and anchor text w}}}}$$

Commonness and keyphraseness



Always the best decision?

Always the best decision?

- This can never help you build an accurate system, because you will always give some wrong links.
- Need to use the context.

Entity Linking - Part II

Pawan Goyal

CSE, IIT Kharagpur

Week 10, Lecture 2

Keyphraseness and Commonness: Always the best decision?

Depth-first search

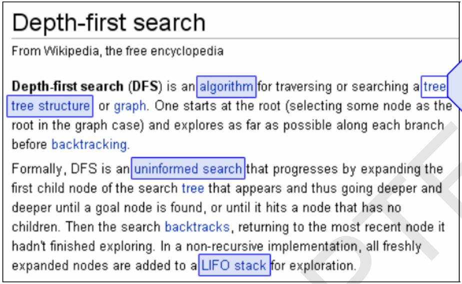
From Wikipedia, the free encyclopedia

Depth-first search (DFS) is an algorithm for traversing or searching a tree tree structure or graph. One starts at the root (selecting some node as the root in the graph case) and explores as far as possible along each branch before backtracking.

Formally, DFS is an uninformed search that progresses by expanding the first child node of the search tree that appears and thus going deeper and deeper until a goal node is found, or until it hits a node that has no children. Then the search backtracks, returning to the most recent node it hadn't finished exploring. In a non-recursive implementation, all freshly expanded nodes are added to a LIFO stack for exploration.

sense	commonness	relatedness
Tree	92.82%	15.97%
Tree (graph theory)	2.94%	59.91%
Tree (data structure)	**2.57%**	**63.26%**
Tree (set theory)	0.15%	34.04%
Phylogenetic tree	0.07%	20.33%
Christmas tree	0.07%	0.0%
Binary tree	0.04%	62.43%
Family tree	0.04%	16.31%
...		

Keyphraseness and Commonness: Always the best decision?



Depth-first search

From Wikipedia, the free encyclopedia

Depth-first search (**DFS**) is an algorithm for traversing or searching a tree, tree structure, or graph. One starts at the root (selecting some node as the root in the graph case) and explores as far as possible along each branch before backtracking.

Formally, DFS is an uninformed search that progresses by expanding the first child node of the search tree that appears and thus going deeper and deeper until a goal node is found, or until it hits a node that has no children. Then the search backtracks, returning to the most recent node it hadn't finished exploring. In a non-recursive implementation, all freshly expanded nodes are added to a LIFO stack for exploration.

sense	commonness	relatedness
Tree	92.82%	15.97%
Tree (graph theory)	2.94%	59.91%
Tree (data structure)	**2.57%**	**63.26%**
Tree (set theory)	0.15%	34.04%
Phylogenetic tree	0.07%	20.33%
Christmas tree	0.07%	0.0%
Binary tree	0.04%	62.43%
Family tree	0.04%	16.31%
...		

Using Relatedness: Basic Idea

- In a sufficiently long text, one finds terms that do not require disambiguation at all.
- Use every unambiguous link in the document as context to disambiguate ambiguous ones.

Computing Relatedness

- Each candidate sense and context term is represented by a single Wikipedia article.

Computing Relatedness

- Each candidate sense and context term is represented by a single Wikipedia article.
- Thus the problem is reduced to selecting the sense article that has most in common with all of the context articles.

Computing Relatedness

- Each candidate sense and context term is represented by a single Wikipedia article.
- Thus the problem is reduced to selecting the sense article that has most in common with all of the context articles.
- Comparison of articles is facilitated by the Wikipedia Link-based measure, which measures the semantic similarity of two Wikipedia pages by comparing their incoming and outgoing links.

Computing Relatedness

- Each candidate sense and context term is represented by a single Wikipedia article.
- Thus the problem is reduced to selecting the sense article that has most in common with all of the context articles.
- Comparison of articles is facilitated by the Wikipedia Link-based measure, which measures the semantic similarity of two Wikipedia pages by comparing their incoming and outgoing links.
- The relatedness of a candidate sense is the weighted average of its relatedness to each context article.

Computing Relatedness

- Each candidate sense and context term is represented by a single Wikipedia article.
- Thus the problem is reduced to selecting the sense article that has most in common with all of the context articles.
- Comparison of articles is facilitated by the Wikipedia Link-based measure, which measures the semantic similarity of two Wikipedia pages by comparing their incoming and outgoing links.
- The relatedness of a candidate sense is the weighted average of its relatedness to each context article.

How to give different weights to the context terms?

Weighting the Context Terms

Weighting the Context Terms

- **link probability:** Use the ones that are almost always used as a link within the articles where they are found, and always link to the same destination

Weighting the Context Terms

- **link probability:** Use the ones that are almost always used as a link within the articles where they are found, and always link to the same destination
- **relatedness:** We can determine how closely a term relates to the central document by calculating its average semantic relatedness to all other context terms

Weighting the Context Terms

- **link probability:** Use the ones that are almost always used as a link within the articles where they are found, and always link to the same destination
- **relatedness:** We can determine how closely a term relates to the central document by calculating its average semantic relatedness to all other context terms

These two variables - link probability and relatedness - are averaged to provide a weight for each context.

Can we improve mention detection with this approach?

- The link detection process starts by gathering all n-grams in the document, and retaining those whose probability exceeds a very low threshold.

Can we improve mention detection with this approach?

- The link detection process starts by gathering all n-grams in the document, and retaining those whose probability exceeds a very low threshold. *Is it the best method?*

Can we improve mention detection with this approach?

- The link detection process starts by gathering all n-grams in the document, and retaining those whose probability exceeds a very low threshold. *Is it the best method?*
- All the remaining phrases are disambiguated using the approach mentioned earlier.

Can we improve mention detection with this approach?

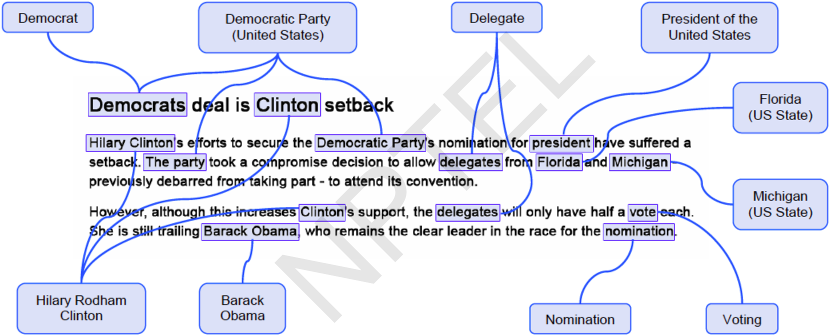
- The link detection process starts by gathering all n-grams in the document, and retaining those whose probability exceeds a very low threshold. *Is it the best method?*
- All the remaining phrases are disambiguated using the approach mentioned earlier.
- This results in a set of associations between terms in the document and the Wikipedia articles that describe them.

Can we improve mention detection with this approach?

- The link detection process starts by gathering all n-grams in the document, and retaining those whose probability exceeds a very low threshold. *Is it the best method?*
- All the remaining phrases are disambiguated using the approach mentioned earlier.
- This results in a set of associations between terms in the document and the Wikipedia articles that describe them.

Can you use this to learn – which concepts should be linked?

Example



Democrat
Democratic Party (United States)
Delegate
President of the United States
Florida (US State)
Michigan (US State)
Democrats deal is Clinton setback
Hilary Clinton's efforts to secure the Democratic Party's nomination for president have suffered a setback. The party took a compromise decision to allow delegates from Florida and Michigan previously debarred from taking part - to attend its convention.
However, although this increases Clinton's support, the delegates will only have half a vote each. She is still trailing Barack Obama, who remains the clear leader in the race for the nomination.
Hilary Rodham Clinton
Barack Obama
Nomination
Voting

The Learning Problem: Which topics should be linked?

- The automatically identified Wikipedia articles provide training instances for a classifier.

The Learning Problem: Which topics should be linked?

- The automatically identified Wikipedia articles provide training instances for a classifier.
- Positive examples are the articles that were manually linked to, while negative ones are those that were not.

The Learning Problem: Which topics should be linked?

- The automatically identified Wikipedia articles provide training instances for a classifier.
- Positive examples are the articles that were manually linked to, while negative ones are those that were not.
- Features of these articles – and the places where they were mentioned – are used to inform the classifier about which topics should and should not be linked.

What are the features?

- **Link Probability:** Average as well as maximum of link probability of the link locations – (e.g. Hillary Clinton and Clinton)
- **Relatedness:** Topics which relate to the central thread of the document are more likely to be linked
- **Disambiguation Confidence:** The confidence score of the classifier for disambiguation
- **Generality:** Defined as the minimum depth at which it is located in Wikipedia's category tree. More useful for the readers to provide links for specific topics.
- **Location and Spread:** Where are these mentioned? First occurrence, last occurrence and the spread.

References

- Mihalcea, Rada, and Andras Csomai. "Wikify!: linking documents to encyclopedic knowledge." Proceedings of the sixteenth ACM conference on information and knowledge management. ACM, 2007.
- Milne, David, and Ian H. Witten. "Learning to link with wikipedia." Proceedings of the 17th ACM conference on Information and knowledge management. ACM, 2008.

Information Extraction - Introduction

Pawan Goyal

CSE, IIT Kharagpur

Week 10, Lecture 3

Goal: "machine reading"

Goal

Acquire structured information knowledge from unstructured text

Information Extraction (IE) Systems

- Find and understand limited relevant parts of texts

Information Extraction (IE) Systems

- Find and understand limited relevant parts of texts
- Gather information from many pieces of text

Information Extraction (IE) Systems

- Find and understand limited relevant parts of texts
- Gather information from many pieces of text
- Produce a structured representation of relevant information:

Information Extraction

Information Extraction (IE) Systems

- Find and understand limited relevant parts of texts
- Gather information from many pieces of text
- Produce a structured representation of relevant information:
  - Relations (in the database sense)
  - A knowledge base

Information Extraction

Information Extraction (IE) Systems

- Find and understand limited relevant parts of texts
- Gather information from many pieces of text
- Produce a structured representation of relevant information:
  - Relations (in the database sense)
  - A knowledge base

Goals

- Organize information so that it is useful to people
- Put information in a semantically precise form that allows further inferences to be made by computer algorithms

Information Extraction (IE)

Definition

Information extraction is the task of finding structured information from unstructured or semi-structured text.

Information Extraction (IE)

Definition

Information extraction is the task of finding structured information from unstructured or semi-structured text.

What sort of information?

Information Extraction (IE)

Definition

Information extraction is the task of finding structured information from unstructured or semi-structured text.

What sort of information?

IE Systems extract clear, factual information

Information Extraction (IE)

Definition

Information extraction is the task of finding structured information from unstructured or semi-structured text.

What sort of information?

IE Systems extract clear, factual information

- Roughly: *Who did what to whom when?* etc.

Information Extraction (IE)

Definition

Information extraction is the task of finding structured information from unstructured or semi-structured text.

What sort of information?

IE Systems extract clear, factual information

- Roughly: *Who did what to whom when?* etc.

E.g., Gathering earnings, profits, headquarters etc. from company reports

- The headquarters of BHP Billiton Limited, and the global headquarters of the combined BHP Billiton Group, are located in Melbourne, Australia.

Information Extraction (IE)

Definition

Information extraction is the task of finding structured information from unstructured or semi-structured text.

What sort of information?

IE Systems extract clear, factual information

- Roughly: *Who did what to whom when?* etc.

E.g., Gathering earnings, profits, headquarters etc. from company reports

- The headquarters of BHP Billiton Limited, and the global headquarters of the combined BHP Billiton Group, are located in Melbourne, Australia.
- headquarters("BHP Biliton Limited", "Melbourne, Australia")

Information Extraction (IE)

Information Extraction (IE)

Example

In 1998, Larry Page and Sergey Brin founded Google Inc.

Information Extraction (IE)

Example

In 1998, Larry Page and Sergey Brin founded Google Inc.

We can extract the following information,

- FounderOf(Larry Page, Google Inc.),
- FounderOf(SergeyBrin, Google Inc.),
- FoundedIn(Google Inc., 1998)

Information Extraction (IE)

Example

In 1998, Larry Page and Sergey Brin founded Google Inc.

We can extract the following information,

- FounderOf(Larry Page, Google Inc.),
- FounderOf(SergeyBrin, Google Inc.),
- FoundedIn(Google Inc., 1998)

Such information can be used by search engines and database management systems to provide better services to end users.

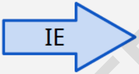
Applications in Biomedical domain

Biomedical domain

- A large amount of scientific publications
- Need to look for discoveries related to particular genes, proteins or other biomedical entities
- Biomedical entities often have synonyms and ambiguous names
- **Critical task:** automatically identify mentions of biomedical entities in text and link them to their corresponding entries in existing knowledge bases.

Biomedical domain

textual abstract:
summary for human



Subject	Relation	Object
p53	is_a	protein
Bax	is_a	protein
p53	has_function	apoptosis
Bax	has_function	induction
apoptosis	involved_in	cell_death
Bax	is_in	mitochondrial outer membrane
Bax	is_in	cytoplasm
apoptosis	related_to	caspase activation
...	...	...

structured knowledge extraction:
summary for machine

Relation Extraction

CHICAGO (AP) — 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, immediately matched the move, spokesman Tim Wagner said. United, a unit of UAL, said the increase took effect Thursday night and applies to most routes where it competes against discount carriers, such as Chicago to Dallas and Atlanta and Denver to San Francisco, Los Angeles and New York.

Subject	Relation	Object
American Airlines	subsidiary	AMR
Tim Wagner	employee	American Airlines
United Airlines	subsidiary	UAL

Relation types

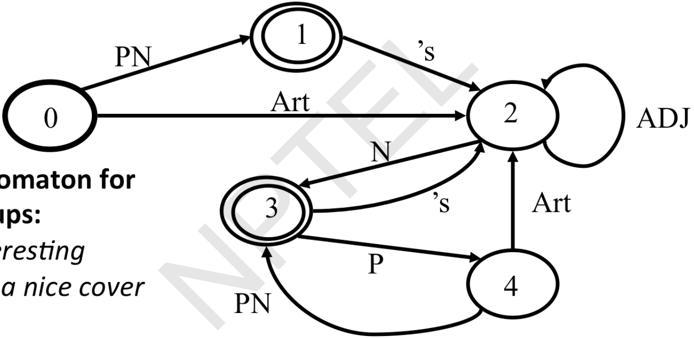
For generic news text ...

Relations		Examples	Types
Affiliations			
	Personal	*married to, mother of*	PER $\rightarrow$ PER
	Organizational	*spokesman for, president of*	PER $\rightarrow$ ORG
	Artifactual	*owns, invented, produces*	(PER $\mid$ ORG) $\rightarrow$ ART
Geospatial			
	Proximity	*near, on outskirts*	LOC $\rightarrow$ LOC
	Directional	*southeast of*	LOC $\rightarrow$ LOC
Part-Of			
	Organizational	*a unit of, parent of*	ORG $\rightarrow$ ORG
	Political	*annexed, acquired*	GPE $\rightarrow$ GPE

Relation extraction: 5 easy methods

- Hand-built patterns
- Bootstrapping methods
- Supervised methods
- Distant supervision
- Unsupervised methods

Hand-written Information Extraction: use regex



Finite Automaton for Noun groups:

John's interesting book with a nice cover

Rule-based Extraction Examples

Determining which person holds what position in what organization

Rule-based Extraction Examples

Determining which person holds what position in what organization

[person], [position] of [org]

Vuk Draskovic, leader of the Serbian Renewal Movement

Rule-based Extraction Examples

Determining which person holds what position in what organization

[person], [position] of [org]

Vuk Draskovic, leader of the Serbian Renewal Movement

[org] (named, appointed, etc.) [person] Prep [office]

NATO appointed Wesley Clark as Commander in Chief

Rule-based Extraction Examples

Determining where an organization is located

Rule-based Extraction Examples

Determining where an organization is located

[org] in [loc]

NATO headquarters in Brussels

Rule-based Extraction Examples

Determining where an organization is located

[org] in [loc]

NATO headquarters in Brussels

[org] [loc] (division, branch, headquarters, etc.)

KFOR Kosovo headquarters

Patterns for learning hyponyms

Intuition from Hearst (1992)

Agar is a substance prepared from a mixture of red algae, such as Gelidium, for laboratory or industrial use.

- What is Gelidium?

Patterns for learning hyponyms

Intuition from Hearst (1992)

Agar is a substance prepared from a mixture of red algae, such as Gelidium, for laboratory or industrial use.

- What is Gelidium?
- How do you know?

Patterns for learning hyponyms

Intuition from Hearst (1992)

Agar is a substance prepared from a mixture of red algae, such as Gelidium, for laboratory or industrial use.

- What is Gelidium?
- How do you know?

Hearst's lexico-syntactic patterns

Automatic Acquisition of Hyponyms

- Y such as $X((, X) * (,$ and/or$) X)$
- such Y as X
- X or other Y
- X and other Y
- Y including X
- Y, especially X

Examples of Hearst patterns

Hearst pattern	Example occurrences
X and other Y	...temples, treasuries, and other important civic buildings.
X or other Y	bruises, wounds, broken bones or other injuries...
Y such as X	The bow lute, such as the Bambara ndang...
such Y as X	...such authors as Herrick, Goldsmith, and Shakespeare.
Y including X	...common-law countries, including Canada and England...
Y, especially X	European countries, especially France, England, and Spain...

Patterns for learning meronyms

Berland and Charniak's patterns

- Selected initial patterns by finding all sentences in a corpus containing *basement* and *building*

Patterns for learning meronyms

Berland and Charniak's patterns

- Selected initial patterns by finding all sentences in a corpus containing *basement* and *building*

whole NN[-PL] 's POS part NN[-PL]	... building's basement ...
part NN[-PL] of PREP {the|a} DET mods [JJ|NN]* whole NN	... basement of a building ...
part NN in PREP {the|a} DET mods [JJ|NN]* whole NN	... basement in a building ...
parts NN-PL of PREP wholes NN-PL	... basements of buildings ...
parts NN-PL in PREP wholes NN-PL	... basements in buildings ...

Problems with hand-built patterns

- Requires hand-building patterns for each relation!
  - hard to write; hard to maintain
  - there are zillions of them
  - domain-dependent
- Don't want to do this for all possible relations!
- Plus, we'd like better accuracy
  - Hearst: 66% accuracy on hyponym extraction
  - Berland and Charniak: 55% accuracy on meronyms

Relation Extraction

Pawan Goyal

CSE, IIT Kharagpur

Week 10, Lecture 4

Bootstrapping approaches

- If you don't have enough annotated text to train on ...
- But you do have:
  - some seed instances of the relation
  - (or some patterns that work pretty well)
  - and lots and lots of unannotated text (e.g., the web)
- can you use those seeds to do something useful?
- Bootstrapping can be considered semi-supervised

Bootstrapping example

- Target relation: burial place

Bootstrapping example

- Target relation: burial place
- Seed tuple : [*Mark Twain, Elmira*]

Bootstrapping example

- Target relation: burial place
- Seed tuple : [*Mark Twain, Elmira*]
- Google for “Mark Twain” and “Elmira”

Bootstrapping example

- Target relation: burial place
- Seed tuple : [*Mark Twain, Elmira*]
- Google for "Mark Twain" and "Elmira"

"Mark Twain is buried in Elmira, NY."
→ X is buried in Y
"The grave of Mark Twain is in Elmira"
→ The grave of X is in Y
"Elmira is Mark Twain's final resting place"
→ Y is X's final resting place

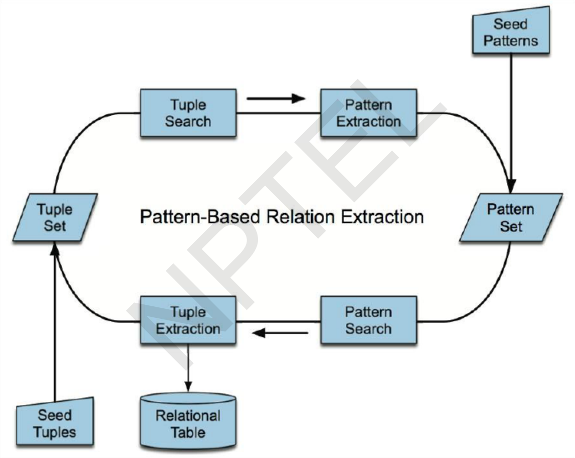
Bootstrapping example

- Target relation: burial place
- Seed tuple : [*Mark Twain, Elmira*]
- Google for “Mark Twain” and “Elmira”

“Mark Twain is buried in Elmira, NY.”
→ X is buried in Y
“The grave of Mark Twain is in Elmira”
→ The grave of X is in Y
“Elmira is Mark Twain’s final resting place”
→ Y is X’s final resting place

- Use those patterns to search for new tuples

Bootstrapping relations



Bootstrapping problems

- Requires that we have seeds for each relation
  - Sensitive to original set of seeds
- Generally have lots of parameters to be tuned
- No probabilistic interpretation
  - Hard to know how confident to be in each result

Supervised Relation Extraction

- Choose a set of relations you would like to extract

Supervised Relation Extraction

- Choose a set of relations you would like to extract
- Find and label data
  - Choose a representative corpus
  - Label the named entities in the corpus
  - Hand-label the relations between these entities
  - Break into training, development and test
- Train a classifier on the training set

Supervised Relation Extraction: An extra step helps

- Find all pairs of named entities (usually in same sentence)

Supervised Relation Extraction: An extra step helps

- Find all pairs of named entities (usually in same sentence)
- **Extra step:** Build a binary classifier to decide if 2 entities are related

Supervised Relation Extraction: An extra step helps

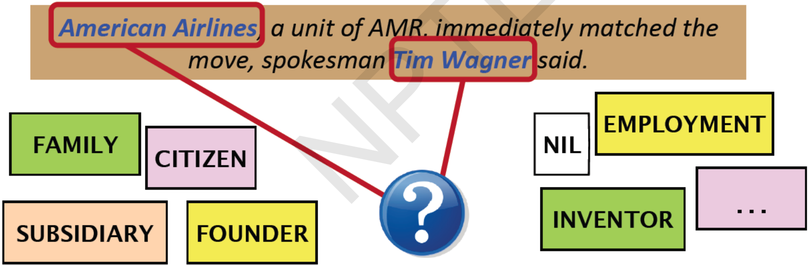
- Find all pairs of named entities (usually in same sentence)
- **Extra step:** Build a binary classifier to decide if 2 entities are related
- If yes, use another classifier to classify the relation

Why the extra step?

- Faster classification training by eliminating most pairs
- Can use distinct feature-sets appropriate for each task

Classify the relation between two entities in a sentence

American Airlines, a unit of AMR, immediately matched the move, spokesman Tim Wagner said



Features: words in mentions $M1$ and $M2$

American Airlines, a unit of AMR, immediately matched the move, spokesman Tim Wagner said.

Features: words in mentions M1 and M2

American Airlines, a unit of AMR, immediately matched the move, spokesman Tim Wagner said.

Bag-of-words features

WM1 = {American, Airlines}, WM2 = {Tim, Wagner}

Features: words in mentions M1 and M2

American Airlines, a unit of AMR, immediately matched the move, spokesman Tim Wagner said.

Bag-of-words features

WM1 = {American, Airlines}, WM2 = {Tim, Wagner}

Head-word features

HM1 = Airlines, HM2 = Wagner, HM12 = Airlines+Wagner

Features: word around the mentions

American Airlines, a unit of AMR, immediately matched the move, spokesman Tim Wagner said.

Features: word around the mentions

American Airlines, a unit of AMR, immediately matched the move, spokesman Tim Wagner said.

Words or bigrams in particular positions left and right of M1/M2

M2:-1 = spokesman, M2: +1 = said

Features: word around the mentions

American Airlines, a unit of AMR, immediately matched the move, spokesman Tim Wagner said.

Words or bigrams in particular positions left and right of M1/M2

M2:-1 = spokesman, M2: +1 = said

Bag of words or bigrams between the two entities

{a, AMR, of, immediately, matched, move, spokesman, the, unit}

Named Entity Type and Mention Level Features

American Airlines, a unit of AMR, immediately matched the move, spokesman Tim Wagner said.

Named Entity Type and Mention Level Features

American Airlines, a unit of AMR, immediately matched the move, spokesman Tim Wagner said.

Named-entity types

M1-NE = ORG, M2-NE = PERSON

Named Entity Type and Mention Level Features

American Airlines, a unit of AMR, immediately matched the move, spokesman Tim Wagner said.

Named-entity types

M1-NE = ORG, M2-NE = PERSON

Concatenation of the two named-entity types

M12-NE = ORG-PERSON

Named Entity Type and Mention Level Features

American Airlines, a unit of AMR, immediately matched the move, spokesman Tim Wagner said.

Named-entity types

M1-NE = ORG, M2-NE = PERSON

Concatenation of the two named-entity types

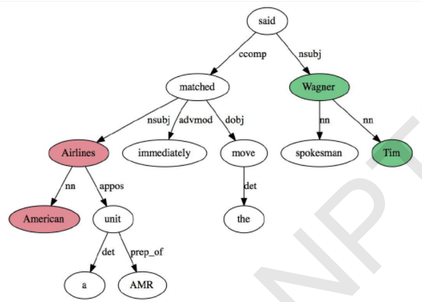
M12-NE = ORG-PERSON

Entity Level of mentions (Name, Nominal, Pronoun)

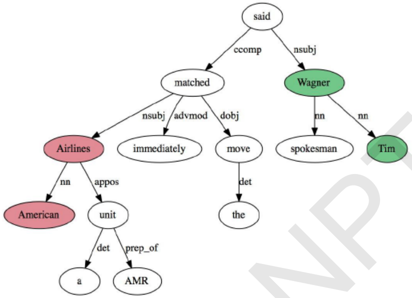
M1:EL = Name, M2:EL = Name

'it' or 'he' would be pronoun, 'the company' would be nominal

Features: dependency syntax features



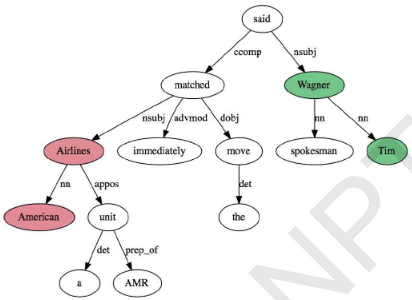
Features: dependency syntax features



Features of mention dependencies

H1DW1 = matched:Airlines
H2DW2 = said:Wagner
Path = {Airlines, matched, said, Wagner}

Features: dependency syntax features



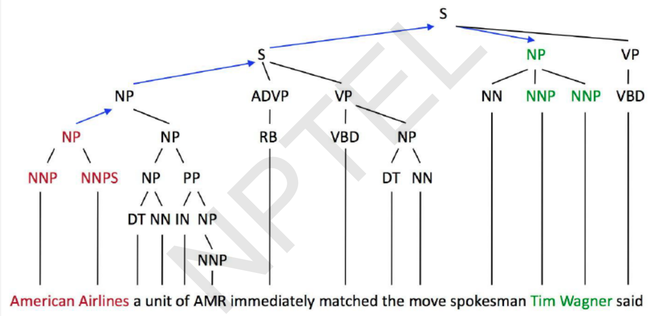
Features of mention dependencies

H1DW1 = matched:Airlines
H2DW2 = said:Wagner
Path = $\{$Airlines, matched, said, Wagner$\}$

Base Phrase Chunk Features

[$_{NP}$ American Airlines], [$_{NP}$ a unit] [$_{PP}$ of] [$_{NP}$ AMR], [$_{ADVP}$ immediately] [$_{VP}$ matched] [$_{NP}$ the move], [$_{NP}$ spokesman Tim Wagner] [$_{VP}$ said].

Features: constituency parse features



Features for relation extraction: Gazetteer and trigger word features

Trigger list for family: kinship terms

Features for relation extraction: Gazetteer and trigger word features

Trigger list for family: kinship terms

parent, wife, husband, grandparent etc. [from Wordnet]

Features for relation extraction: Gazetteer and trigger word features

Trigger list for family: kinship terms

parent, wife, husband, grandparent etc. [from Wordnet]

Gazetteer

List of useful geo or geopolitical words

- Country name list
- Other sub-entries

Relation extraction classifiers

Now you can use any classifier

- SVM
- MaxEnt (multiclass logistic regression)
- Naïve Bayes
- etc.

Relation extraction classifiers

Now you can use any classifier

- SVM
- MaxEnt (multiclass logistic regression)
- Naïve Bayes
- etc.

Train it on the training set, tune on the development set, test on the test set

Evaluation of Supervised Relation Extraction

Compute $P/R/F_1$ for each relation

$$P = \frac{\text{Number of correctly extracted relations}}{\text{Total number of extracted relations}}$$

Evaluation of Supervised Relation Extraction

Compute $P/R/F_1$ for each relation

$$P = \frac{\text{Number of correctly extracted relations}}{\text{Total number of extracted relations}}$$

$$R = \frac{\text{Number of correctly extracted relations}}{\text{Total number of gold relations}}$$

Evaluation of Supervised Relation Extraction

Compute $P/R/F_1$ for each relation

$$P = \frac{\text{Number of correctly extracted relations}}{\text{Total number of extracted relations}}$$

$$R = \frac{\text{Number of correctly extracted relations}}{\text{Total number of gold relations}}$$

$$F_1 = \frac{2PR}{P+R}$$

Supervised RE : summary

Supervised approach can achieve high accuracy

- At least, for some relations
- If we have lots of hand-labeled training data

But has significant limitations!

- Labeling large training set (+ named entities) is expensive
- Doesn't generalize to different relations

Supervised RE : summary

Supervised approach can achieve high accuracy

- At least, for some relations
- If we have lots of hand-labeled training data

But has significant limitations!

- Labeling large training set (+ named entities) is expensive
- Doesn't generalize to different relations

Beyond supervised relation extraction

- Distantly supervised relation extraction
- Unsupervised relation extraction

Distant Supervision

Pawan Goyal

CSE, IIT Kharagpur

Week 10, Lecture 5

Distant supervision paradigm

Hypothesis

If two entities belong to a certain relation, any sentence containing those two entities is likely to express that relation

Distant supervision paradigm

Hypothesis

If two entities belong to a certain relation, any sentence containing those two entities is likely to express that relation

Key Idea

Use a database of relations to get lots of training examples

- instead of hand-crafting a few seed tuples (bootstrapping)
- instead of using hand-labeled corpus (supervised)

Distant supervision paradigm

Hypothesis

If two entities belong to a certain relation, any sentence containing those two entities is likely to express that relation

Key Idea

Use a database of relations to get lots of training examples

- instead of hand-crafting a few seed tuples (bootstrapping)
- instead of using hand-labeled corpus (supervised)

Approach

For each pair of entities in a large database:

- Grab sentences containing these entities from a corpus
- Extract lots of noisy features from the sentences
  - Lexical features, syntactic features, named entity tags
- Combine in a classifier

Benifits of distant supervision

Has advantages of supervised approach

- leverage rich, reliable hand-crafted knowledge
- relations have canonical names
- can use rich features (e.g. syntactic features)

Benifits of distant supervision

Has advantages of supervised approach

- leverage rich, reliable hand-crafted knowledge
- relations have canonical names
- can use rich features (e.g. syntactic features)

Has advantages of unsupervised approach

- leverage unlimited amounts of text data
- allows for very large number of weak features
- not sensitive to training corpus: genre independent

Hypernyms via distant supervision

Construct a noisy training set consisting of occurrences from a corpus, that contain hyponym-hypernym pair from Wordnet.

Ex: Shakespeare - author

Hypernyms via distant supervision

Construct a noisy training set consisting of occurrences from a corpus, that contain hyponym-hypernym pair from Wordnet.
Ex: Shakespeare - author

Training yields high-signal examples like:

- "...consider authors like Shakespeare..."
- "Some authors (including Shakespeare)..."
- "Shakespeare was the author of several..."
- "Shakespeare, author of The Tempest..."

Hypernyms via distant supervision

Construct a noisy training set consisting of occurrences from a corpus, that contain hyponym-hypernym pair from Wordnet.
Ex: Shakespeare - author

Training yields high-signal examples like:

- "...consider authors like Shakespeare..."
- "Some authors (including Shakespeare)..."
- "Shakespeare was the author of several..."
- "Shakespeare, author of The Tempest..."

But also noisy examples like:

- "The author of Shakespeare in Love..."
- "...authors at the Shakespeare Festival..."

Learning hypernym patterns

- Take corpus sentence
  ... doubly heavy hydrogen atom called deuterium ...

Learning hypernym patterns

- Take corpus sentence
  ... doubly heavy hydrogen atom called deuterium ...
- Collect noun pairs
  e.g. (atom, deuterium)
  752,311 pairs from 6M sentences of newswire

Learning hypernym patterns

- Take corpus sentence
  ... doubly heavy hydrogen atom called deuterium ...
- Collect noun pairs
  e.g. (atom, deuterium)
  752,311 pairs from 6M sentences of newswire
- Is pair an IS-A in WordNet?
  14, 387 yes; 737, 924 no

Learning hypernym patterns

- Take corpus sentence
  ... doubly heavy hydrogen atom called deuterium ...
- Collect noun pairs
  e.g. (atom, deuterium)
  752,311 pairs from 6M sentences of newswire
- Is pair an IS-A in WordNet?
  14, 387 yes; 737, 924 no
- Parse the sentences
- Extract patterns
- Train classifer on patterns
  logistic regression with 70K features

Syntactic dependency paths

Patterns are based on paths through dependency parses generated by MINIPAR.

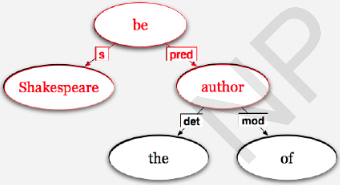
Syntactic dependency paths

Patterns are based on paths through dependency parses generated by MINIPAR. Example word pair: (Shakespeare, author)
Example sentence: "Shakespeare was the author of several plays..."

Syntactic dependency paths

Patterns are based on paths through dependency parses generated by MINIPAR. Example word pair: (Shakespeare, author)
Example sentence: “Shakespeare was the author of several plays...”

Minipar parse:

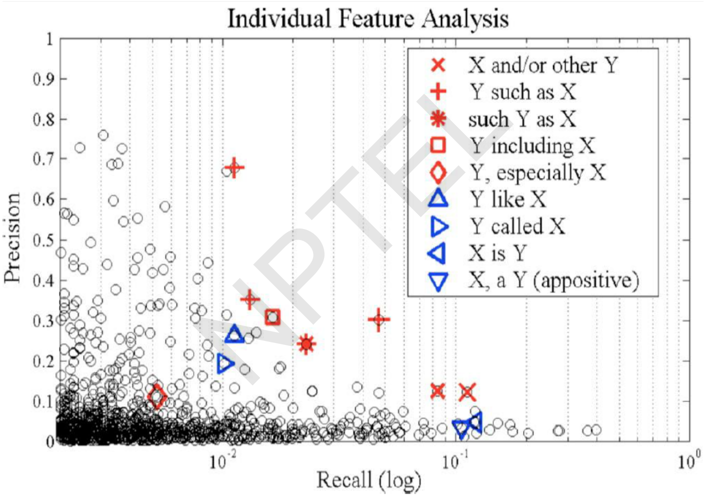


Extract shortest path:
-N:s:VBE, be, VBE:pred:N

Syntactic dependency paths

- Original nouns in the noun pair are removed to create a more general pattern
- Each dependency path is presented as on ordered list of dependency tuples
- Optional “satellite links” are added to each shortest path
  “such NP as NP” : function word ‘such’ is added to the shortest dependency path

Precision-Recall of hypernym extraction patterns



What about other relations

Mintz, Bills, Snow, Jurafsky (2009).
Distant supervision for relation extraction without labeled data.

Training set	**Corpus**
Freebase	Wikipedia
102 relations 940,000 entities 1.8 million instances	1.8 million articles 25.7 million sentences

Fequent Freebase relations

Relation name	Size	Example
/people/person/nationality	281,107	John Dugard, South Africa
/location/location/contains	253,223	Belgium, Nijlen
/people/person/profession	208,888	Dusa McDuff, Mathematician
/people/person/place_of_birth	105,799	Edwin Hubble, Marshfield
/dining/restaurant/cuisine	86,213	MacAyo's Mexican Kitchen, Mexican
/business/business_chain/location	66,529	Apple Inc., Apple Inc., South Park, NC
/biology/organism_classification_rank	42,806	Scorpaeniformes, Order
/film/film/genre	40,658	Where the Sidewalk Ends, Film noir
/film/film/language	31,103	Enter the Phoenix, Cantonese
/biology/organism_higher_classification	30,052	Calopteryx, Calopterygidae
/film/film/country	27,217	Turtle Diary, United States
/film/writer/film	23,856	Irving Shulman, Rebel Without a Cause
/film/director/film	23,539	Michael Mann, Collateral
/film/producer/film	22,079	Diane Eskenazi, Aladdin
/people/deceased_person/place_of_death	18,814	John W. Kern, Asheville
/music/artist/origin	18,619	The Octopus Project, Austin
/people/person/religion	17,582	Joseph Chartrand, Catholicism
/book/author/works_written	17,278	Paul Auster, Travels in the Scriptorium
/soccer/football_position/players	17,244	Midfielder, Chen Tao
/people/deceased_person/cause_of_death	16,709	Richard Daintree, Tuberculosis
/book/book/genre	16,431	Pony Soldiers, Science fiction
/film/film/music	14,070	Stavisky, Stephen Sondheim
/business/company/industry	13,805	ATS Medical, Health care

Collecting training data

Corpus text

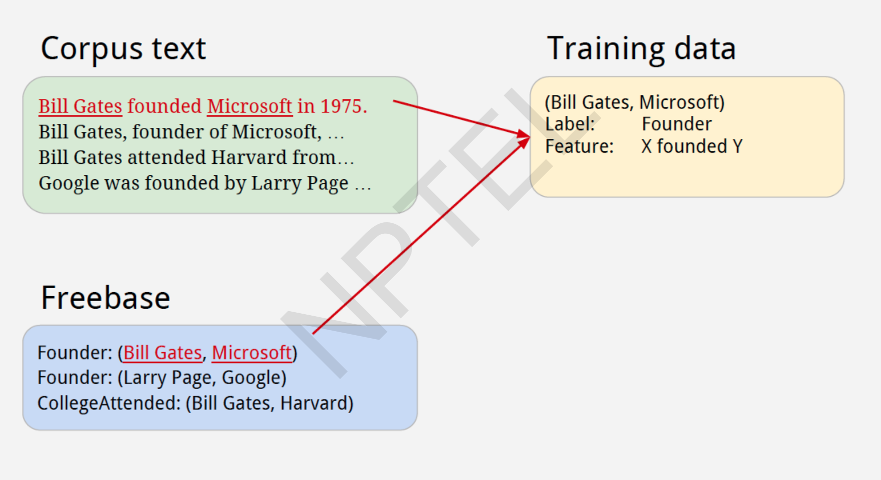
Bill Gates founded Microsoft in 1975.
Bill Gates, founder of Microsoft, ...
Bill Gates attended Harvard from...
Google was founded by Larry Page ...

Training data

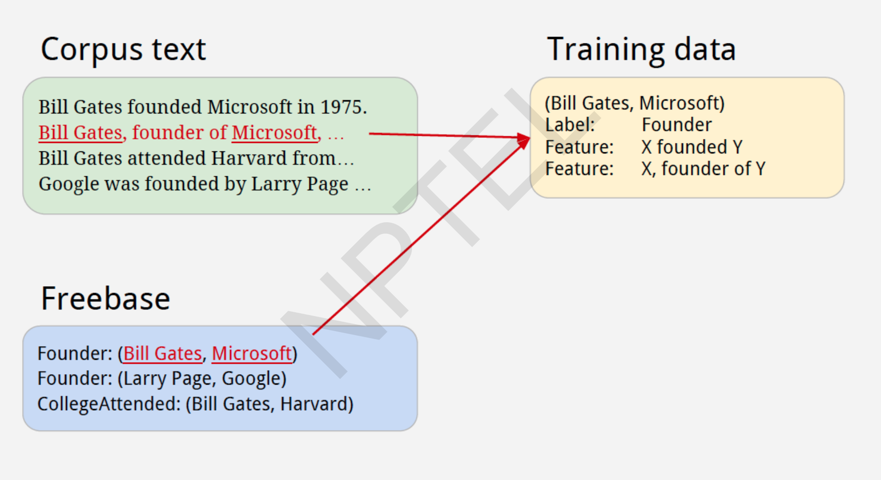
Freebase

Founder: (Bill Gates, Microsoft)
Founder: (Larry Page, Google)
CollegeAttended: (Bill Gates, Harvard)

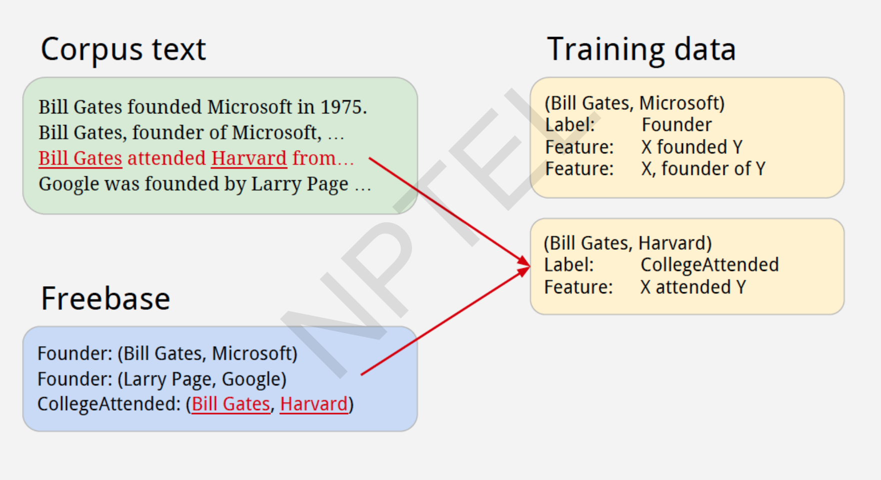
Collecting training data



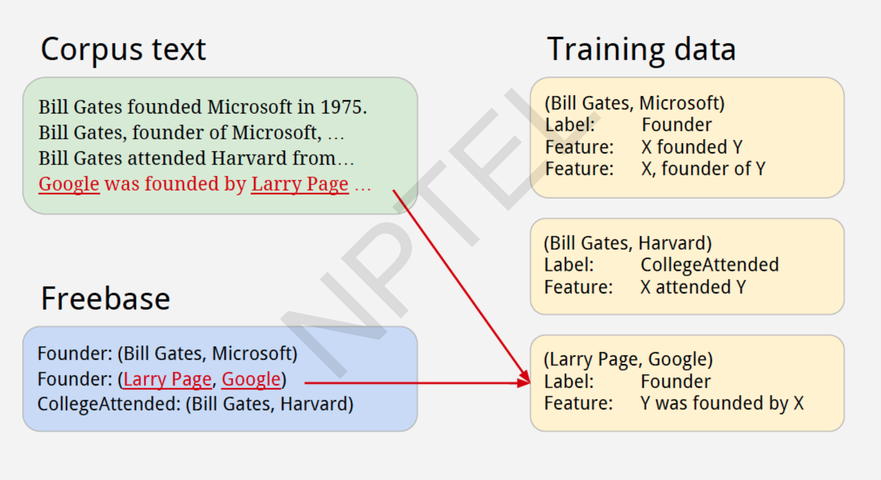
Collecting training data



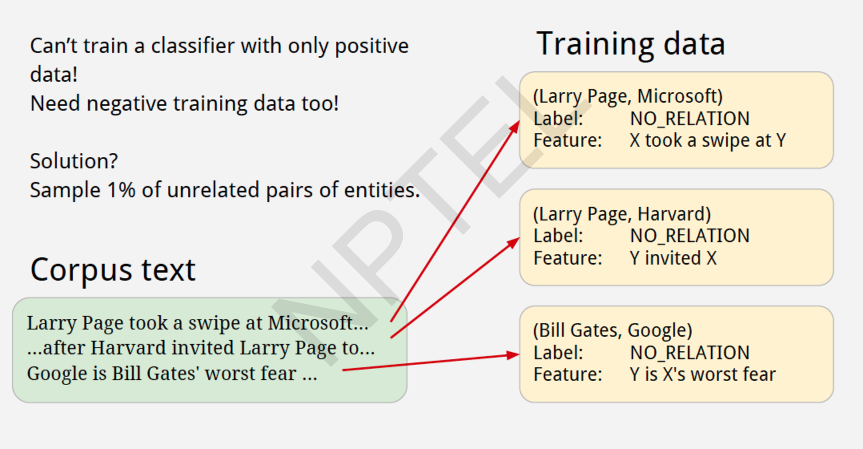
Collecting training data



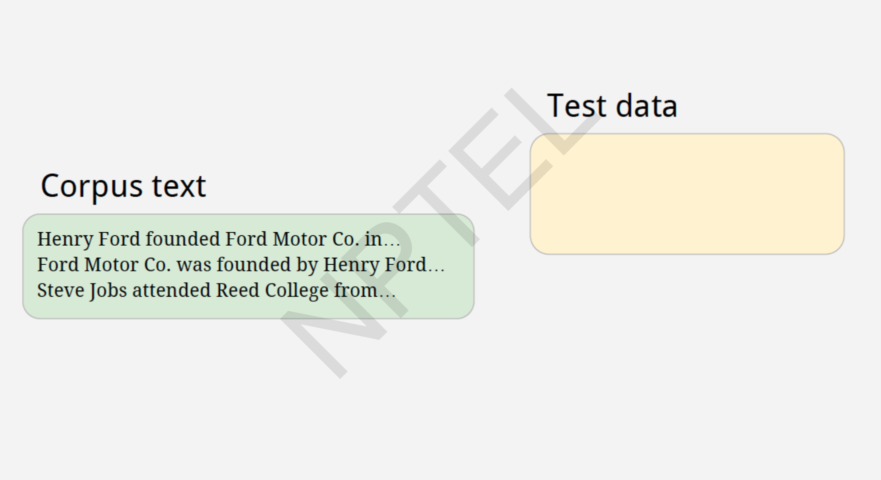
Collecting training data



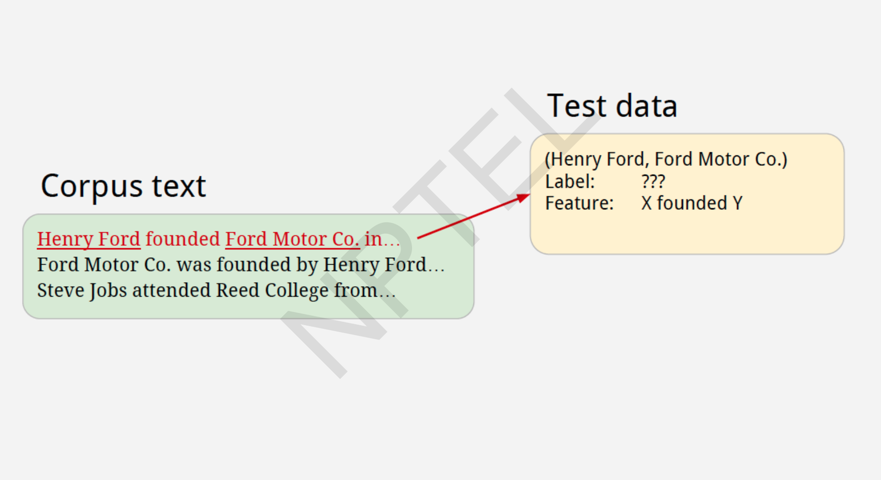
Negative training data

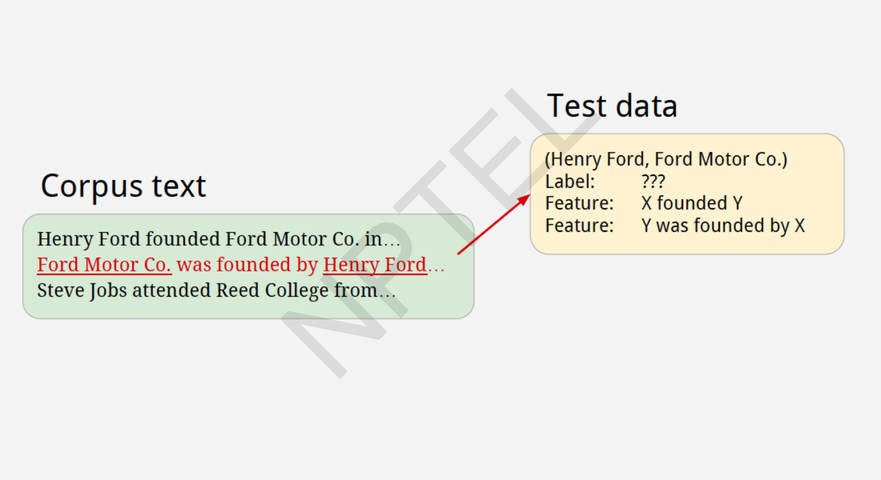


Preparing test data

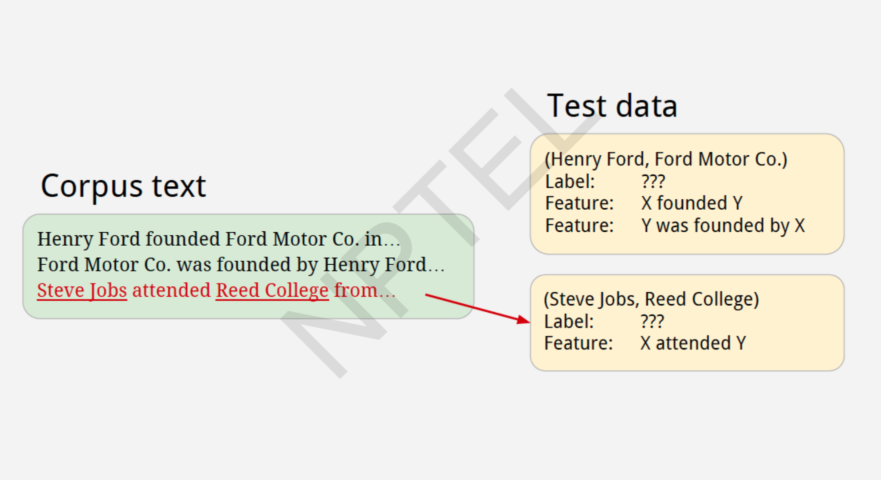


Preparing test data

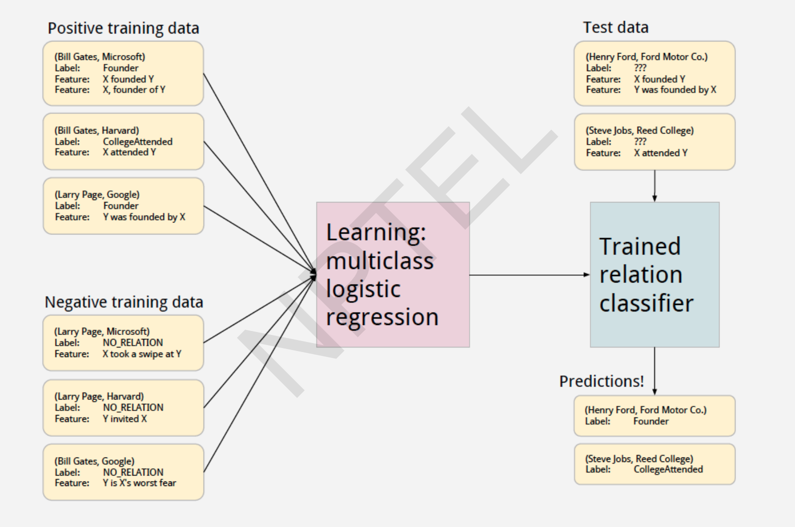



Corpus text
Henry Ford founded Ford Motor Co. in…
Ford Motor Co. was founded by Henry Ford…
Steve Jobs attended Reed College from…
Test data
(Henry Ford, Ford Motor Co.)
Label: ???
Feature: X founded Y
Feature: Y was founded by X

Preparing test data



The experiment



Features

Each feature describes how two entities are related in a sentence, using either syntactic or non-syntactic information.

Lexical Features

- The sequence of words between the two entities
- The POS tags of these words
- A window of k words to the left of Entity 1 and their POS tags
- A window of k words to the right of Entity 2 and their POS tags

Feature conjunction

- Each lexical feature consists of the conjunction of all these components
- A conjuctive feature is generated for each $k \in \{0, 1, 2\}$