Information Extraction

Announcements

- Course evaluation: please fill out
- Bring laptop on Wednesday
  - NLP research at Columbia
  - Final review
- Final exam: 12/9, final is cumulative
- What topics would you like to hear about again?

What topics would you like covered in review?

Summarization

Dialog

Machine translation

Neural net architectures

Attention

Word embeddings

Word sense disambiguation

Semantics

Are there any other topics you would like covered?

Summing up on bias.

- There are methods for de-biasing word embeddings
  - Remove gender bias by zeroing the gender project for each word on a pre-defined gender direction (Boukbasi et al 2016)
  - Train debiased word embeddings from scratch by concentrating gender information in last coordinate – encourage words in different groups to differ in last coordinate

loss function.

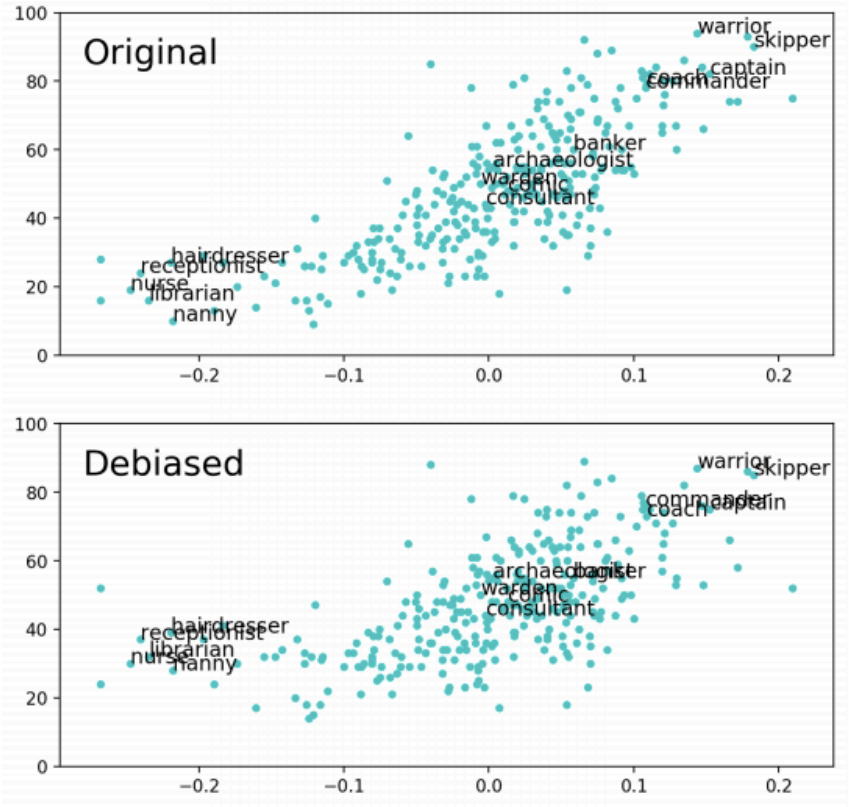
Do debiasing methods work?

- Lipstick on a pig paper claims they do not
- Hides the bias

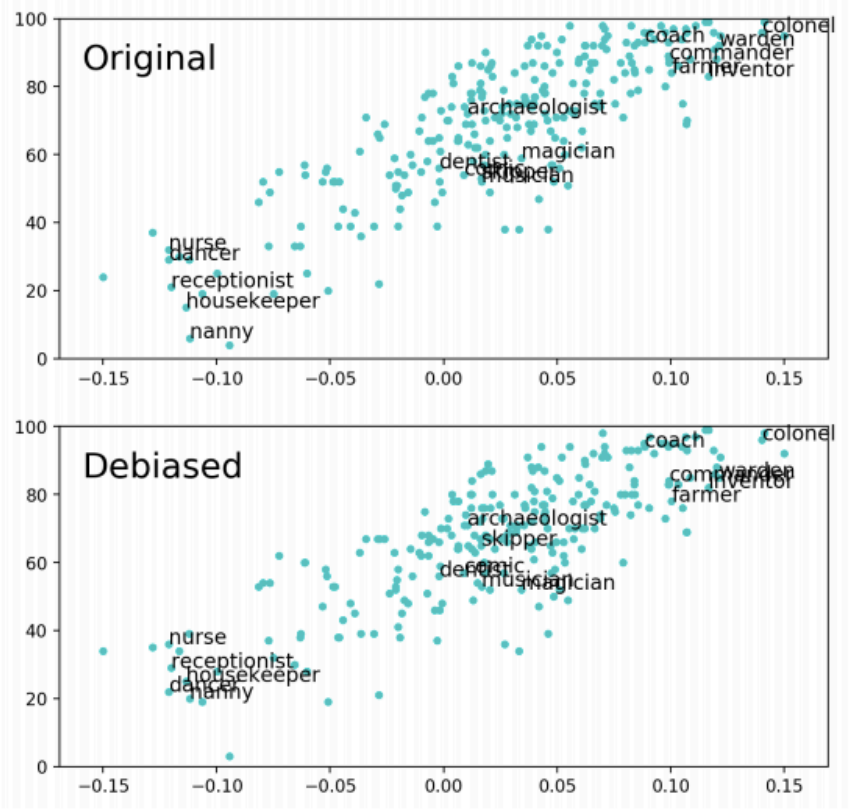
- Still reflected in similarities between gender neutral words
  - E.g., “math” “delicate”

- Three experiments 2000 words from early paper claimed gender bias.
  - Do male and female biased words cluster together?
  - % male/female socially biased words among k nearest neighbors of target word
  - Predict the gender of a word

GloVe, train unbiased/biased. 差别只有92%与95% accuracy



(a) The plots for HARD-DEBIASED embedding, before (top) and after (bottom) debiasing.



(b) The plots for GN-GLOVE embedding, before (top) and after (bottom) debiasing.

Do we care if embeddings are biased?

Information Extraction

- Extraction of concrete facts from text

anything with capital letters

- Named entities, relations, events

- Often used to create a structured knowledge base of facts

eg., google try to answer questions from unstructured text

- Kathy McKeown, a professor from Columbia University in New York City, took a train yesterday to Washington DC.

Named Entities

- Kathy McKeown$_{per}$, a professor from Columbia University$_{org}$ in New York City$_{loc}$, took a train yesterday to Washington DC$_{loc}$.

Named Entities, Relations

- Kathy McKeown$_{per}$, a professor from Columbia University$_{org}$ in New York City$_{loc}$, took a train yesterday to Washington DC$_{loc}$.

relation: between entities.

- Kathy McKeown from Columbia
- Columbia in New York City

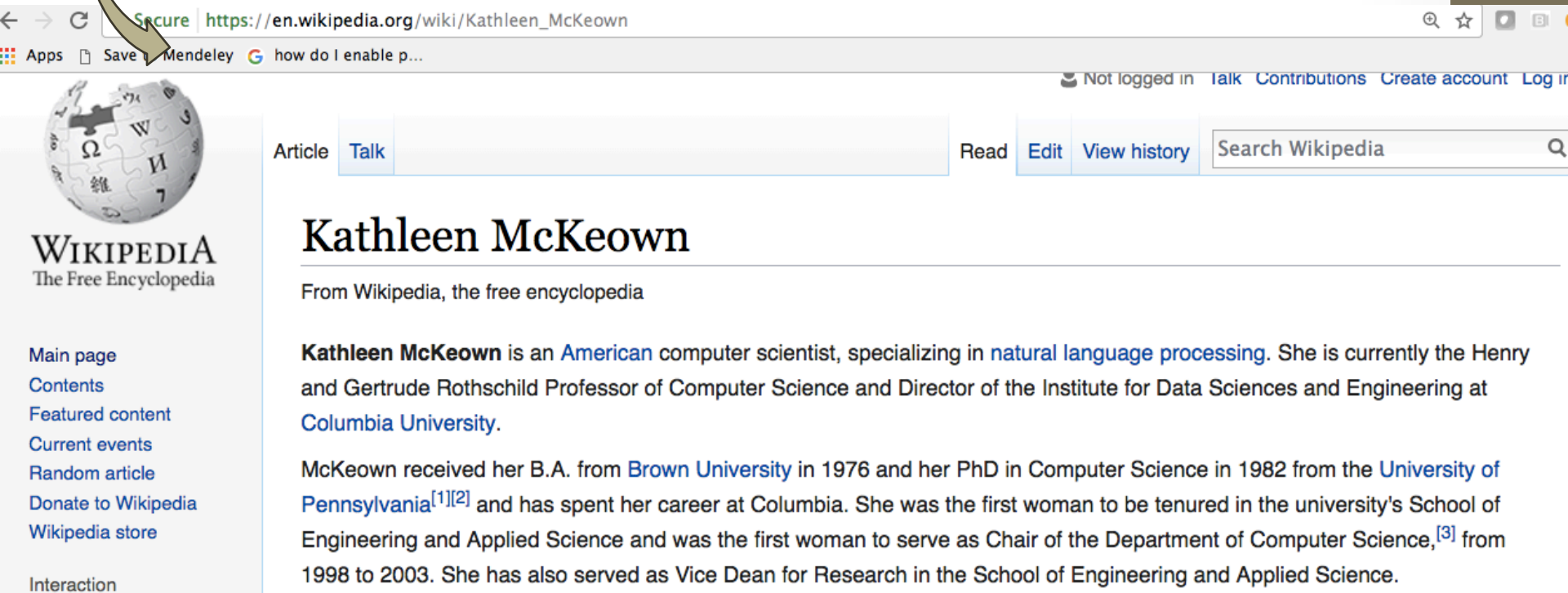
Named Entities, Relations, Events

- Kathy McKeown$_{per}$, a professor from Columbia University$_{org}$ in New York City$_{loc}$, took a train yesterday to Washington DC$_{loc}$.

- Kathy McKeown took a train (yesterday)^{time}

Entity Discovery and Linking

Kathy McKeown, a professor from Columbia University in New York City, took a train yesterday to Washington DC.



IE for Template Filling Relation Detection

Given a set of documents and a domain of interest, fill a table of required fields.

- For example:

  Number of car accidents per vehicle type and number of casualties in the accidents.

Vehicle Type	# accidents	# casualties	Weather
SUV	1200	190	Rainy
Trucks	200	20	Sunny

Never-Ending Language Learner

Tom Mitchell
CMU

- Can computers learn to read?
- Browses the web and attempts to extract facts from hundreds of millions of web pages
- Attempts to improve its methods and accuracy
- To date, 50 million candidate facts at different levels of confidence
- http://rtw.ml.cmu.edu/rtw/

IE for Question Answering

Q: When was Gandhi born?
A: October 2, 1869

Q: Where was Bill Clinton educated?
A: Georgetown University in Washington, D.C.

Q: What was the education of Yassir Arafat?
A: Civil Engineering

Q: What is the religion of Noam Chomsky?
A: Jewish

State of the Art (English)

	F-measure
• Named Entities (news)	• 89%
• Relations (slot filling)	• 59%
• Events (nuggets)	• 63%

以前用了很久，类似于POS　　跟HMM有关

Methods: Sequence labeling (MEMM, CRF), neural nets, distant learning

现在NN比较多　　hard to get labeled data.

比如noun, verb, etc

Features: linguistic features, similarity, popularity, gazeteers, ontologies, verb triggers

knowledge based

Where Have You Been Entity Discovery and Linking?

Grow with DEFT	2006-2011	2012-2017 *HENG JI, RPI*
Mention Extraction	Human (most)	Automatic
NIL Clustering	None	64 methods
Foreign Languages	Chinese (5%-10% lower than English)	**System for 282 languages (Chinese/Spanish comparable to/Outperform English); research toward 3,000 languages**
Document Size	-	500 →90,000 documents
Genre	News, web blog	**News, Discussion Forum, Web blog, Tweets**
Entity Types	PER, GPE, ORG GPE, 比如united states	**PER, GPE, ORG, LOC, FAC, hundreds of fine-grained types for typing**
Mention Types	Name or all concepts (most)	Name, Nominal, Pronoun (for BeST) Nominal, professor
KB	Wikipedia	Freebase → List only
Training Data	20,000 queries (entity mentions)	**500 → 0 documents; unsupervised linking comparable to supervised linking**
#(Good) Papers	62	110 (new KBP track at ACL); 6 tutorials at top conferences

Slide from Heng Ji

Approach for NER

- <PERSON>**Alexander Mackenzie**</PERSON> , (<TIMEX >**January 28, 1822** <TIMEX> - <TIMEX>**April 17, 1892**</TIMEX>), **a building contractor and writer, was the second** Prime Minister of <GPE>**Canada**</GPE> **from ….**

- **Statistical sequence labeling techniques can be used – similar to POS tagging**
  - **Word-by-word sequence labeling**
  - **Example of features**
    - **POS tags**
    - **Syntactic constituents**
    - **Shape features**
    - **Presence in a named entity list**

Supervised Approach for relation detection

- Given a corpus of annotated relations between entities, train a classifier:
  - A binary classifier
    - Given a span of text and two entities -> decide if there is a relationship between these two entities
- Features
  - Types of two named entities
  - Bag of words
  - POS of words in between
- Example:
  - A rented SUV went out of control on Sunday, causing the death of seven people in Brooklyn
  - Relation: Type = Accident, Vehicle Type = SUV, casualty = 7, weather = ?

relation不见得特别明显，还有可能有很多词，所以蛮难的

Pattern Matching for Relation Detection

- Patterns:
  - "[CAR_TYPE] went out of control on [TIMEX], causing the death of [NUM] people"
  - "[PERSON] was born in [GPE]"
  - "[PERSON] was graduated from [FAC]"
  - "[PERSON] was killed by <X>"
- **Matching Techniques**
  - **Exact matching**
    - Pros and Cons?
  - **Flexible matching (e.g., [X] was .* killed .* by [Y])**
    - Pros and Cons?

exact matching 需要太多list.
但还在使用：certain structured genres, high precision though low recall.

Is rule-based exact matching still used (take a guess)?

Yes

No

What problems would arise with flexible matching?

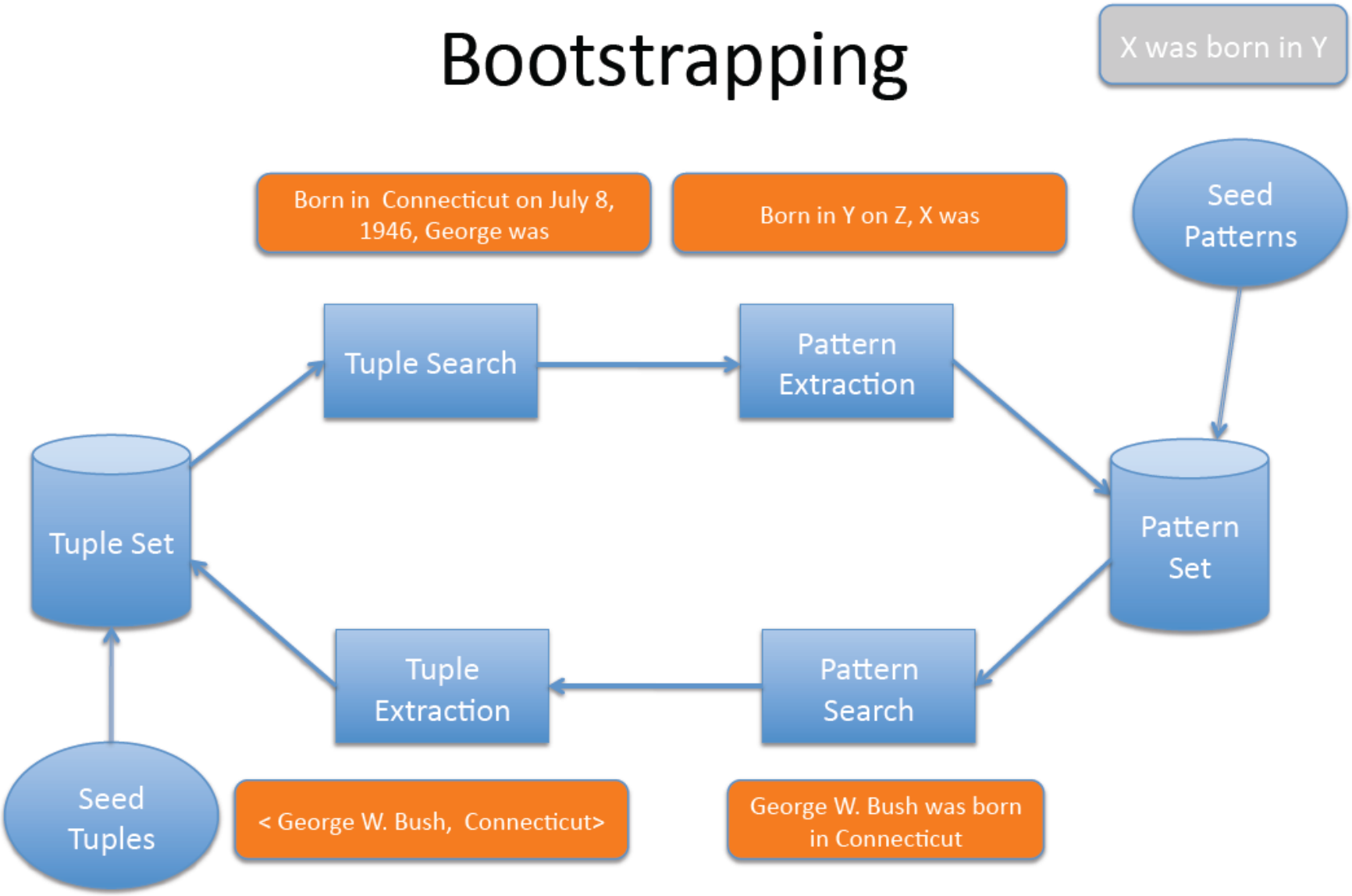
low precision, still need a lot of rules.

Pattern Matching

- How can we come up with these patterns?
- Manually?
  - Task and domain-specific
  - Tedious, time consuming, not scalable
- Machine learning, semi-supervised approaches

start with a few patterns.

Bootstrapping



Task:
Produce a biography of [person]

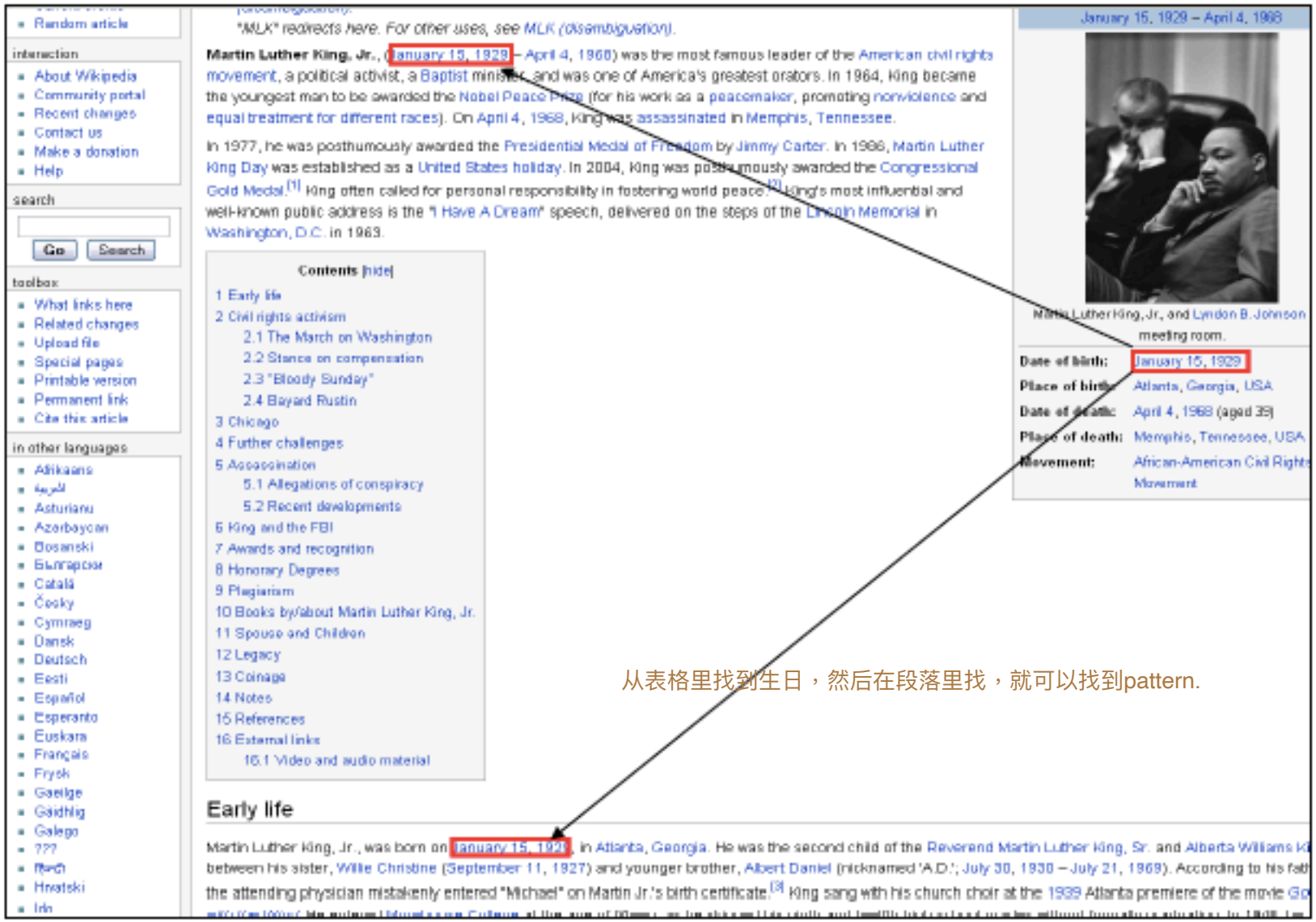
1. Name(s), aliases:
2. *Date of Birth or Current Age:
3. *Date of Death:
4. *Place of Birth:
5. *Place of Death:
6. Cause of Death:
7. Religion (Affiliations):
8. Known locations and dates:
9. Last known address:
10. Previous domiciles:
11. Ethnic or tribal affiliations:
12. Immediate family members
13. Native Language spoken:
14. Secondary Languages spoken:
15. Physical Characteristics
16. Passport number and country of issue:
17. Professional positions:
18. Education
19. Party or other organization affiliations:
20. Publications (titles and dates):

Biography – two approaches

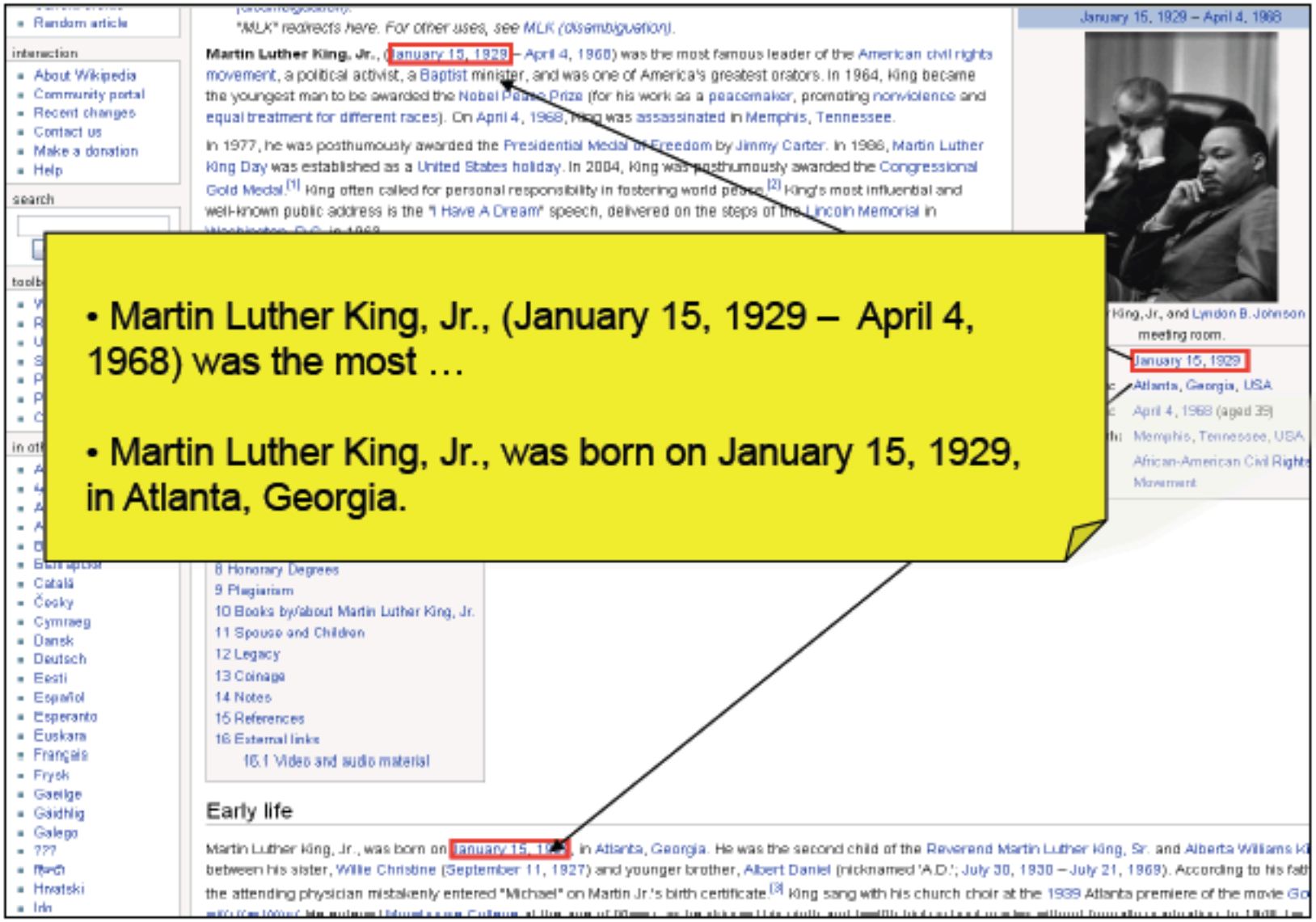
- To obtain high precision, handle each slot independently using bootstrapping to learn IE patterns.

- To improve the recall, utilize a biographical sentence classifier

Biography patterns from Wikipedia



Biography patterns from Wikipedia



Run NER on these sentences

- <Person> Martin Luther King, Jr. </Person>, (<Date>January 15, 1929</Date> – <Date> April 4, 1968</Date>) was the most…
- <Person> Martin Luther King, Jr. </Person>, was born on <Date> January 15, 1929 </Date>, in <GPE> Atlanta, Georgia </GPE>.
- Take the token sequence that includes the tags of interest + some context (2 tokens before and 2 tokens after)

Convert to Patterns:

- <Target_Person> (<Target_Date> – <Date>) was the
- <Target_Person> , was born on <Target_Date>, in
- Remove more specific patterns – if there is a pattern that contains other, take the smallest > *k* tokens.
- ➔ <Target_Person> , was born on <Target_Date>
- ➔ <Target_Person> (<Target_Date> – <Date>)
- Finally, verify the patterns manually to remove irrelevant patterns.

Examples of Patterns:

- 502 distinct place-of-birth patterns:
  - 600 <Target_Person> was born in <Target_GPE>
  - 169 <Target_Person> (born <Date> in <Target_GPE>)
  - 44 Born in <Target_GPE> , <Target_Person>
  - 10 <Target_Person> was a native <Target_GPE>
  - 10 <Target_Person> 's hometown of <Target_GPE>
  - 1 <Target_Person> was baptized in <Target_GPE>
  - ...
- 291 distinct date-of-death patterns:
  - 770 <Target_Person> (<Date> - <Target_Date>)
  - 92 <Target_Person> died on <Target_Date>
  - 19 <Target_Person> <Date> - <Target_Date>
  - 16 <Target_Person> died in <GPE> on <Target_Date>
  - 3 < Target_Person> passed away on < Target_Date >
  - 1 < Target_Person> committed suicide on <Target_Date>
  - ...

Biography as an IE task

- This approach is good for the consistently annotated fields in Wikipedia: *place of birth, date of birth, place of death, date of death*

- Not all fields of interests are annotated, a different approach is needed to cover the rest of the slots

Bouncing between Wikipedia and Google

- Use **one** seed tuple **only:**
  - <Target Person> and <Target field>
    - Google: "Arafat" "civil engineering", we get:

只有一个training example. 称为few shot learning.

Few shot learning

Web

Yasser **Arafat**
By 1956 **Arafat** graduated with a bachelor's degree in **civil engineering** and served as a second lieutenant in the Egyptian Army during the Suez Crisis. ...
www.jewishvirtuallibrary.org/jsource/biography/**arafat**.html - 61k -
Cached - Similar pages - Note this

Yasser **Arafat**: Biography and Much More from Answers.com
In the 1950s, **Arafat** studied at Fu'ad I University in Cairo (now Cairo University), majoring in **civil engineering**. He was reportedly a member of the Muslim ...
www.answers.com/topic/yasser-**arafat** - 89k - Cached - Similar pages - Note this

Engology.com, Engineer Yasser **Arafat**, Nobel Piece Prize Winner ...
After the war **Arafat** studied **civil engineering** at the University of Cairo. He headed the Palestinian Students League and, by the time he graduated, ...
www.engology.com/engpg5eyasser**arafat**.htm - 7k - Cached - Similar pages - Note this

Yasser **Arafat** and the Palestine Liberation Organization
It was there that Yasser **Arafat**, a **Civil Engineering** student, and his coterie, including Salah Khalaf (Abu Iyad), later to become **Arafat**'s second in command ...
www.palestinefacts.org/pf_1948to1967_plo_**arafat**.php - 14k -
Cached - Similar pages - Note this

A Life in Retrospect: Yasser **Arafat** | TIME
Here's one thing we know for sure: Yasser **Arafat** was a grand ... at King Fuad I University (now Cairo University), where he studied **civil engineering**. ...
www.time.com/time/world/article/0,8599,781566-1,00.html - 39k -
Cached - Similar pages - Note this

Yassir **Arafat's** Biography
Yasser **Arafat** was born in 1929 in Jerusalem. His full name is: Mohammed Abad Arouf **Arafat**. He studied **civil engineering** at CairoUniversity. ...
www.eretzyisroel.org/~jkatz/**arafat**bio.html - 72k - Cached - Similar pages - Note this

Biographical and other information on Yasser **Arafat** who is in bad ...
In 1951, at the age of 21, **Arafat** got military training with the Egyptian army. — In 1956, **Arafat** earned a degree in **civil engineering** at the University of...
www.freemuslims.org/news/article.php?article=198 - 14k -
Cached - Similar pages - Note this

Bouncing between Wikipedia and Google

- Use one seed tuple only:
  - Google: "Arafat" "civil engineering", we get:
    - ⇒ **Arafat** *graduated with a bachelor's degree* in **civil engineering**
    - ⇒ **Arafat** *studied* **civil engineering**
    - ⇒ **Arafat**, *a* **civil engineering** *student*
    - ⇒ ...
  - Using these snippets, corresponding patterns are created, then filtered out.

from one seed, get multiple patterns.

Bouncing between Wikipedia and Google

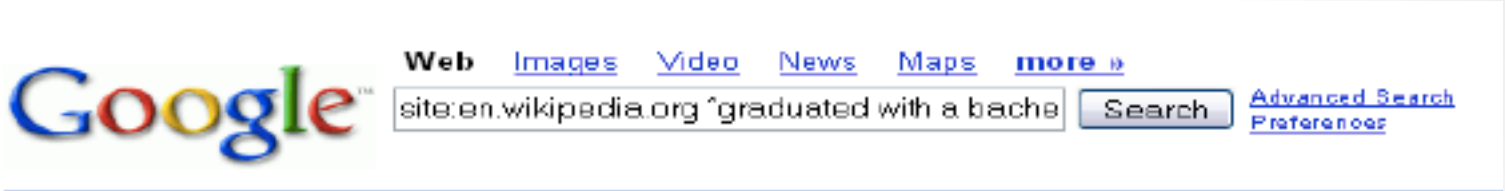
- Use one seed tuple only:
  - Google: "Arafat" "civil engineering", we get:
    - ⇒ **Arafat** ***graduated with a bachelor's degree*** in **civil engineering**
    - ⇒ **Arafat** *studied* **civil engineering**
    - ⇒ **Arafat**, *a* **civil engineering** *student*
    - ⇒ ...
  - Using these snippets, corresponding patterns are created, then filtered out manually
  - Due to time limitation the automatic filter was not completed.

- To get more seed tuples, go to Wikipedia biography pages only and search for:
  - ***"graduated with a bachelor's degree in"***
  - We get:

Bouncing between Wikipedia and Google

- **New seed tuples:**
  - “Burnie Thompson” “political science“
  - “Henrey Luke” “Environment Studies”
  - “Erin Crocker” “industrial and management engineering”
  - “Denise Bode” “political science”
  - …
- Go back to Google and repeat the process to get more seed patterns!

这种是semi-supervised learning.



Web Resu

[Burnie Thompson - Wikipedia, the free encyclopedia]
In 2000, he **graduated with a bachelor's degree in** political science from California State University, Fullerton. Two years later he graduated from The ...
en.wikipedia.org/wiki/Burnie_Thompson - 19k - Cached - Similar pages - Note this

[Roscoe Lee Browne - Wikipedia, the free encyclopedia]
Born in Woodbury, New Jersey, Browne first attended historically black Lincoln University in Pennsylvania, and **graduated with a bachelor's degree in** 1946. ...
en.wikipedia.org/wiki/Roscoe_Lee_Browne - 38k - Cached - Similar pages - Note this

[Henry Luke Orombi - Wikipedia, the free encyclopedia]
Robert has **graduated with a Bachelor's Degree in** Environment Studies from Makerere University and Daniel, a gifted musician like his father, is working on ...
en.wikipedia.org/wiki/Henry_Luke_Orombi - 25k - Cached - Similar pages - Note this

[Gustave Eiffel - Wikipedia, the free encyclopedia]
Eiffel's study habits improved and he **graduated with a bachelor's degree in** both science and humanities. Eiffel went on to attend college at Sainte Barbe ...
en.wikipedia.org/wiki/Gustave_Eiffel - 52k - Cached - Similar pages - Note this

[Erin Crocker - Wikipedia, the free encyclopedia]
... New York, where she **graduated with a bachelor's degree in** industrial and management engineering in 2003. In 2002, Crocker signed with Woodring Racing to ...
en.wikipedia.org/wiki/Erin_Crocker - 30k - Cached - Similar pages - Note this

[Jim Boeheim - Wikipedia, the free encyclopedia]
Boeheim enrolled in Syracuse University as a student in 1963 and **graduated with a bachelor's degree in** social science in 1969(SU Athletics). ...
en.wikipedia.org/wiki/Jim_Boeheim - 30k - Cached - Similar pages - Note this

[Denise Bode - Wikipedia, the free encyclopedia]
She **graduated with a bachelor's degree in** political science from the University of Oklahoma where she chaired the University of Oklahoma Student Congress. ...

Bouncing back and forth

- Worked well for a fields such as education, publications, immediate family members, party, other organization activities
- Did not work well for other fields including religion, ethnic or tribal affiliations, previous domiciles -> too much noise
- Why is the bouncing idea better than using only one corpus?

Why is bouncing back between two sources better than using just one?

variety&diversity (google > wikipedia)
structure&accuracy->wikipedia

How are neural nets used for IE?

Organizing knowledge

It's a version of ***Chicago*** – the standard classic Macintosh menu font, with that distinctive thick diagonal in the "N".	***Chicago*** was used by default for Mac menus through MacOS 7.6, and OS 8 was released mid-1997..	***Chicago VIII*** was one of the early 70s-era ***Chicago*** albums to catch my ear, along with ***Chicago II***.

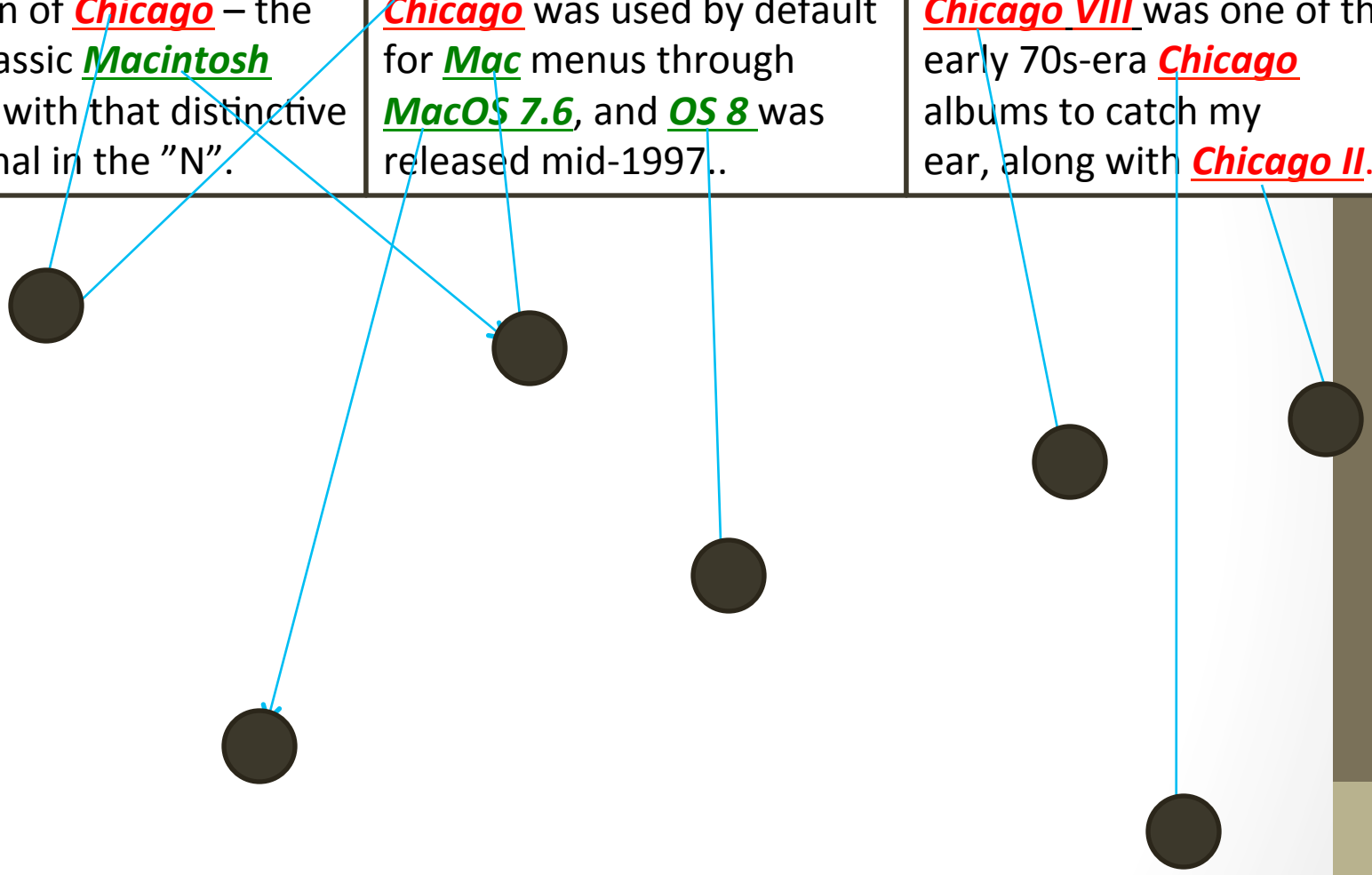
cross entity linking

-> meaning of Chicago

Slide from Heng Ji

Cross-document co-reference resolution

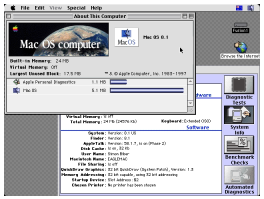
It's a version of ***Chicago*** – the standard classic ***Macintosh*** menu font, with that distinctive thick diagonal in the "N".	***Chicago*** was used by default for ***Mac*** menus through ***MacOS 7.6***, and ***OS 8*** was released mid-1997..	***Chicago VIII*** was one of the early 70s-era ***Chicago*** albums to catch my ear, along with ***Chicago II***.

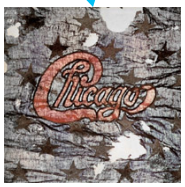


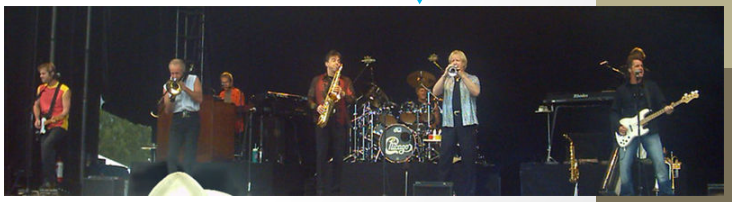
Slide from Heng Ji

Reference resolution: (disambiguation to Wikipedia)

It's a version of ***Chicago*** – the standard classic ***Macintosh*** menu font, with that distinctive thick diagonal in the "N".	***Chicago*** was used by default for ***Mac*** menus through ***MacOS 7.6***, and ***OS 8*** was released mid-1997..	***Chicago VIII*** was one of the early 70s-era ***Chicago*** albums to catch my ear, along with ***Chicago II***.







Slide from Heng Ji

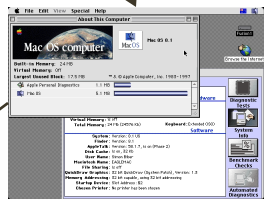
The "Reference" Collection has Structure

It's a version of ***Chicago*** – the standard classic ***Macintosh*** menu font, with that distinctive thick diagonal in the "N".	***Chicago*** was used by default for ***Mac*** menus through ***MacOS 7.6***, and ***OS 8*** was released mid-1997..	***Chicago VIII*** was one of the early 70s-era ***Chicago*** albums to catch my ear, along with ***Chicago II***.

Used_In

Is_a

Is_a



Succeeded

Released

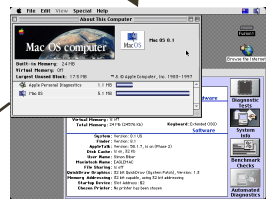
Slide from Heng Ji

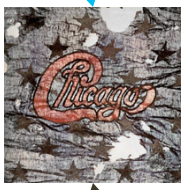
Analysis of Information Networks

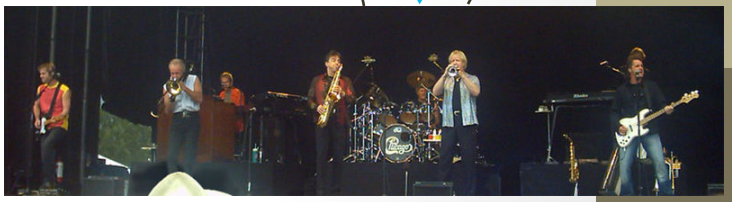
It's a version of ***Chicago*** – the standard classic ***Macintosh*** menu font, with that distinctive thick diagonal in the "N".

Chicago was used by default for ***Mac*** menus through ***MacOS 7.6***, and ***OS 8*** was released mid-1997..

Chicago VIII was one of the early 70s-era ***Chicago*** albums to catch my ear, along with ***Chicago II***.







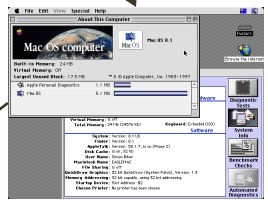
Slide from Heng Ji

Here – Wikipedia as a knowledge resource .... bu can use other resources

Used_In

Is_a

Is_a



Succeeded

Released

Slide from Heng Ji

Wikification: The Reference Problem

Cycles of Knowledge: Grounding for/ using Knowledge

Blumenthal (D) is a candidate for the U.S. Senate seat now held by Christopher Dodd (D), and he has held a commanding lead in the race since he entered it. But the Times report has the potential to fundamentally reshape the contest in the Nutmeg State.

Richard Blumenthal
From Wikipedia, the free encyclopedia

Democratic Party (United States)
From Wikipedia, the free encyclopedia

United States Senate
From Wikipedia, the free encyclopedia

Blumenthal (D) is a candidate for the U.S. Senate seat now held by Christopher Dodd (D), and he has held a commanding lead in the race since he entered it. But the Times report has the potential to fundamentally reshape the contest in the Nutmeg State.

Chris Dodd
From Wikipedia, the free encyclopedia

The New York Times
From Wikipedia, the free encyclopedia

Connecticut
From Wikipedia, the free encyclopedia

Slide from Heng Ji

Task Definition

- A formal definition of the task consists of:

1. A definition of the **mentions** (concepts, entities) to highlight

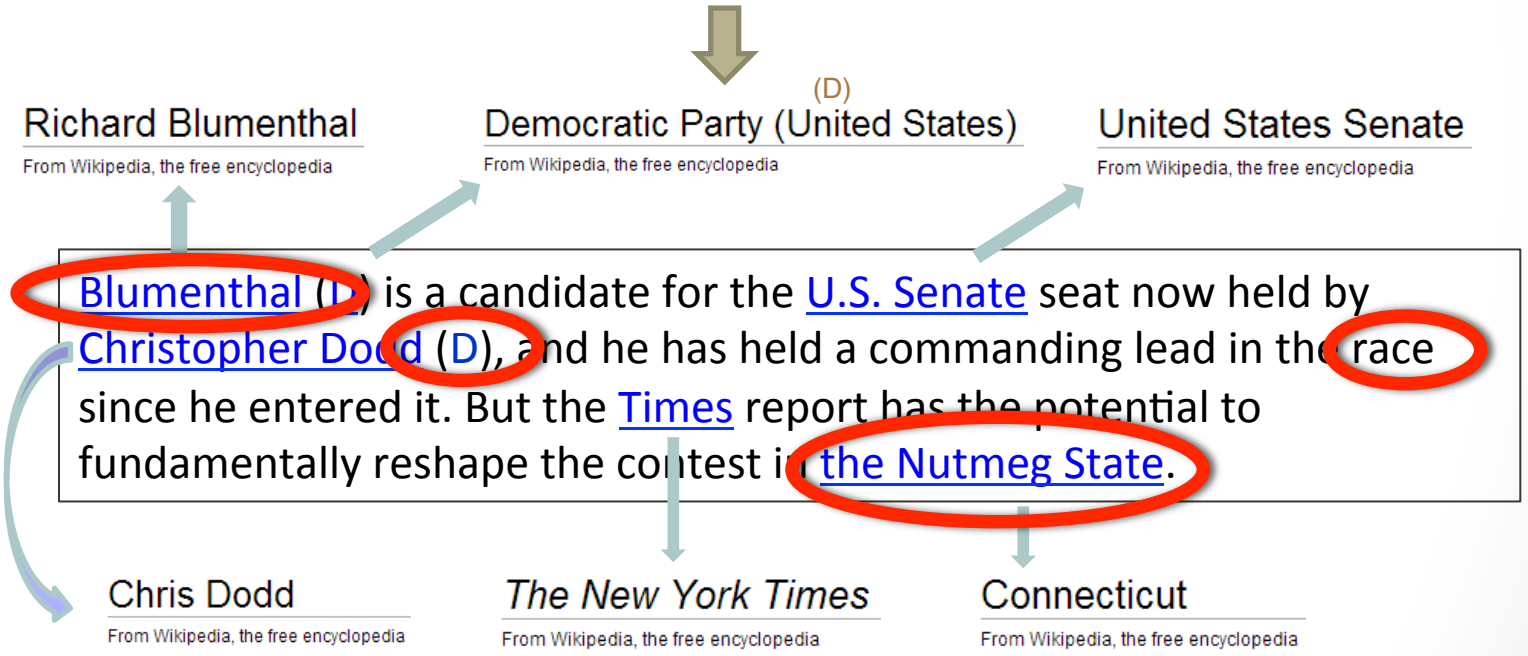
2. Determining the target encyclopedic resource (**KB**)

3. Defining what to point to in the KB (**title**)

Slide from Heng Ji

Examples of Mentions (1)

Blumenthal (D) is a candidate for the U.S. Senate seat now held by Christopher Dodd (D), and he has held a commanding lead in the race since he entered it. But the Times report has the potential to fundamentally reshape the contest in the Nutmeg State.



Slide from Heng Ji

Neural Approach to Entity Linking (Wikification)

Gupta, Singh and Roth, EMNLP 2017

- Learns a dense, unified representation of entities
  - Encodes semantic and background knowledge from multiple sources
  - An encoder for each source of information
  - Entity embeddings learned to be similar to encodings
- Only uses indirect supervision from Wikipedia/Freebase
- Can incorporate new entities without retraining existing representations
- http://cogcomp.org/papers/GuptaSiRo17.pdf

Jointly Embedding Entity Information

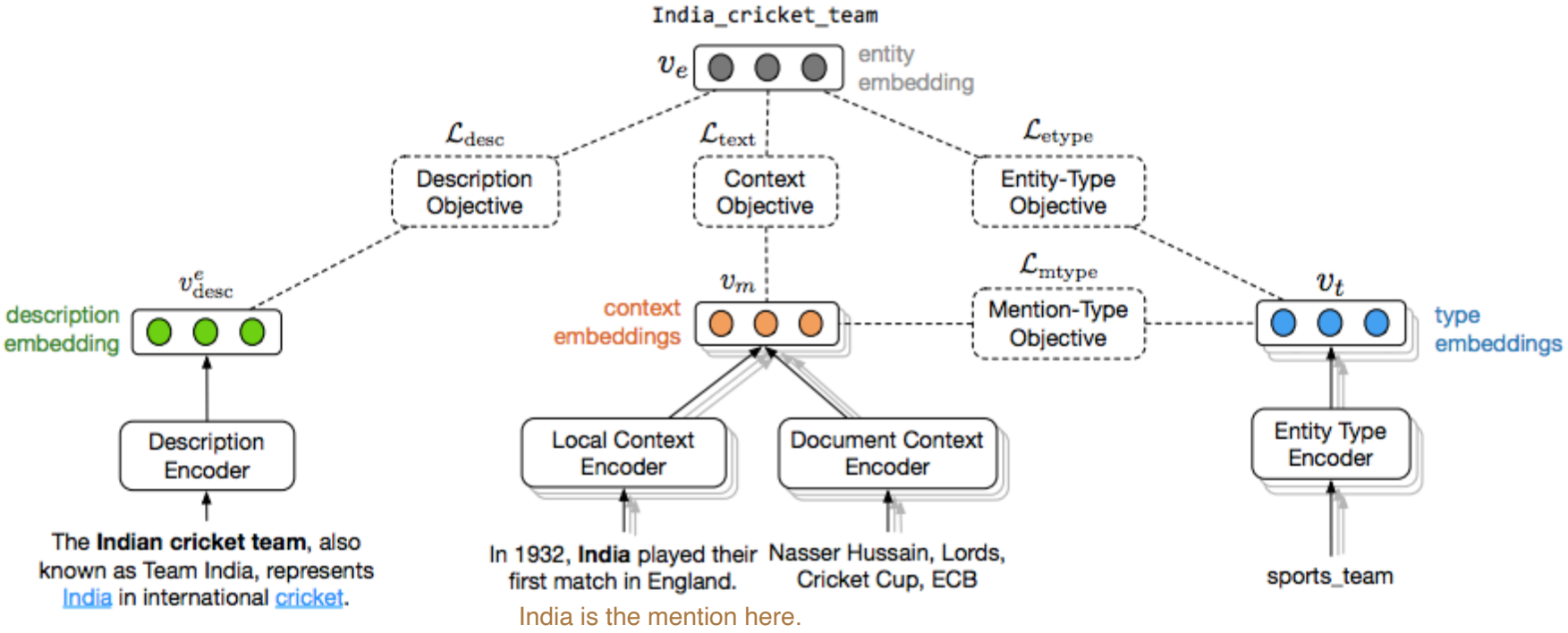


Figure 1: **Overview of the Model** (§ 3): Each entity has a Wikipedia description, linked mentions in Wikipedia (only one shown), and fine-grained types from Freebase (only one shown). We encode local and document-level mention contexts (§ 3.1), entity-description (§ 3.2), and fine-grained entity-types (§ 3.3 & § 3.4). Joint optimization (§ 3.5) over these provides the unified entity representations $\{v_e\}$.

Jointly Embedding Entity Information

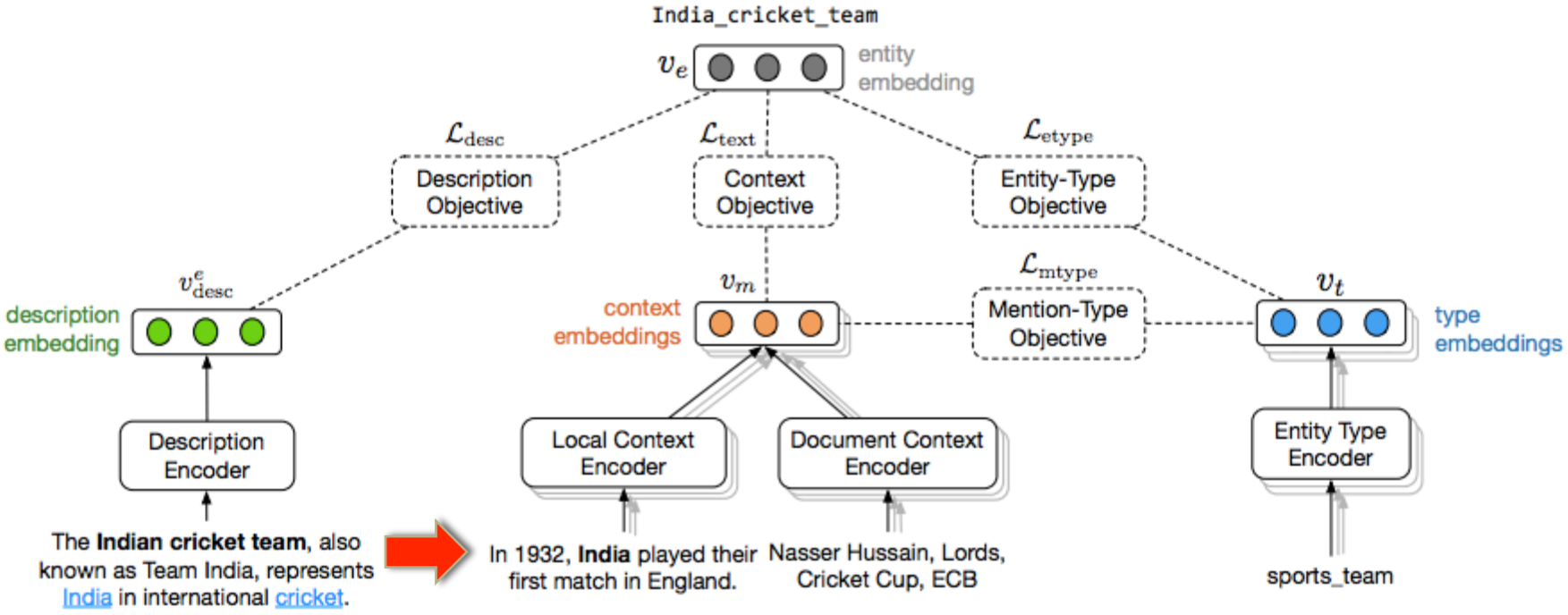


Figure 1: **Overview of the Model** (§ 3): Each entity has a Wikipedia description, linked mentions in Wikipedia (only one shown), and fine-grained types from Freebase (only one shown). We encode local and document-level mention contexts (§ 3.1), entity-description (§ 3.2), and fine-grained entity-types (§ 3.3 & § 3.4). Joint optimization (§ 3.5) over these provides the unified entity representations $\{v_e\}$.

Look at Wikipedia

- Entity description: https://en.wikipedia.org/wiki/India_national_cricket_team

Encoding the mention context

In 1932, India played their first game in England.

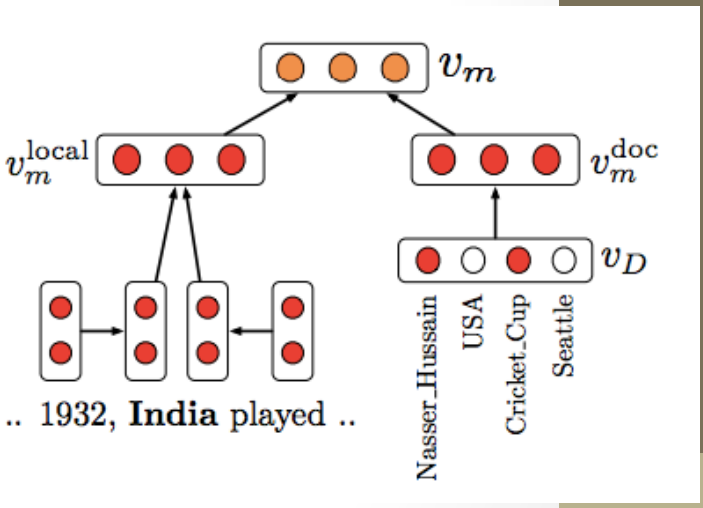
- Example mention contains two mentions: "*India*" and "*England*"
- Aim to disambiguate "*India*" to the team
  - **Local context**: "*played*" and "*match*"
  - **Document context**: to identify the sport
- Preserve the semantics: "*England*" should not match to a team

Local Context

- Given mention m in sentence: {w_1, …, m, …. w_N}
- Left LSTM applied to w_1…m -> $\overrightarrow{h^l_m}$
- Right LSTM applied to m …. w_N -> $\overleftarrow{h^r_m}$
- $\overrightarrow{h^l_m}; \overleftarrow{h^r_m}$ concatenated and passed through a single layer feed forward network

Document Context Encoder

- Bag of mentions vector:
  - USA, Pearl Jam, Nasser Hassain
- Compressed to a low dimensional representation using a single layer feed forward neural network
- Combine local and document representations to get a mention level encoding using concatenation and feed through a single layer feed forward network



$$v_m \in \mathbb{R}^d.$$

Joint inference using all mentions

As of 7 January 2019, *India* have played 533 Test matches

As of 16 June 2018, India have played 968 ODI matches

- Use an objective that encourages the mentions v_m to be similar to the entity description v_e
  - Maximize the probability of predicting the correct entity from the mentions

$$\mathcal{L}_{\text{text}} = \frac{1}{M} \sum_{i=1}^{M} \log P_{\text{text}}(e_{m^{(i)}} | m^{(i)})$$

Encoding Entity Description D

- Embed each word of the Wikipedia description as a d-dimensional vector

- Encode as a fixed vector using a CNN:

$$v^e_{\text{desc}} \in \mathbb{R}^d$$

Learning the Type Representation

- Embed type T in Freebase
- Each entity can have multiple types
- Jointly learn entity and type representations

Learning Unified Entity Representations

- Separate models for entity mentions, entity descriptions, type descriptions

- To learn the different entity representations and their parameters, jointly maximize the total objective

$$\{v_e\}, \Theta = \underset{\{v_e\},\Theta}{\operatorname{argmax}} \mathcal{L}_{\text{text}} + \mathcal{L}_{\text{desc}} + \mathcal{L}_{\text{etype}} + \mathcal{L}_{\text{mtype}}$$

where v_e are the set of entity representations and θ are the parameters

	CoNLL		ACE05	Wiki
	Test	Dev		
Plato (Sup)	79.7	-	-	-
Plato (Semi-Sup)	86.4	-	-	-
*AIDA**	*81.8*	-	-	-
*BerkCNN:Sparse**	*74.9*	-	*83.6*	*81.5*
*BerkCNN:CNN**	*81.2*	*86.91*	*84.5*	*75.7*
*BerkCNN:Full**	*85.5*	-	*89.9*	*82.2*
Priors	68.5	70.9	81.1	78.1
Model C	81.4	83.4	83.7	86.1
Model CD	81.0	83.2	85.8	86.1
Model CT	82.3	83.9	86.5	88.2
Model CDT	82.5	85.6	86.8	88.0
Model CDTE	82.9	84.9	85.6	89.0

Table 1: **Entity Linking Performance:** Accuracy

	F1	Accuracy
AIDA	77.8	-
Wikifier	85.1	-
Vinculum	88.5	-
Model C	88.9	93.1
Model CDT	89.8	93.9
Model CDTE	90.7	94.3

Table 2: **Results for ACE-2004**: F1 is calculated for predicted mentions, and accuracy on gold-mentions. Results for Wikifier and AIDA are from (Ling et al., 2015). All systems use the same mention extraction protocol showing the difference in F1 is due to linking performance.

Looking forward

- More languages: 3000!
- Multi-media
- Streaming mode
- No more training data
- Context-aware, living