Entities

what are they,
where are they
and
how do we model them?

Daniel M. Bikel
Google Research
TAC 2020
2021-02-22

Google Research

About the speaker (who am I?)

Name: Dan Bikel (rhymes with “Brickell”)

Been at Google Research since 2010,
    … minus 2.25 years leading NLP team at LinkedIn

Some past stuff

- Nymble: HMM-based named entity recognizer (BBN)
- Syntactic parser (UPenn)
- IE, QA, everything (IBM)
- Semantic parser for Google Now → Assistant (Google)
- Applications of NLP, MT and deep learning methods to profiles and feed (LinkedIn)

Currently lead project called EKG on specialized domain understanding

Most recent focus: entity discovery (joint work with Andrew McCallum and others)

About this talk

What is an entity, anyway?

How and where can we find entities? Two angles:

- automatic KG induction
- entities in multiple languages

What can we do with entities?

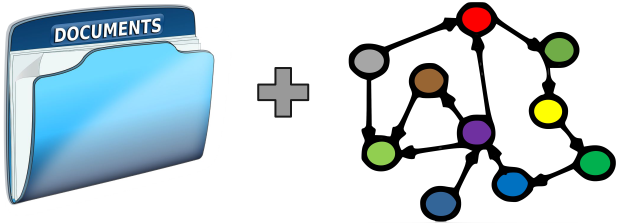
- entity linking without alias tables
- incorporation into LM’s and NLU systems
- new approaches to entity discovery

What is an entity?

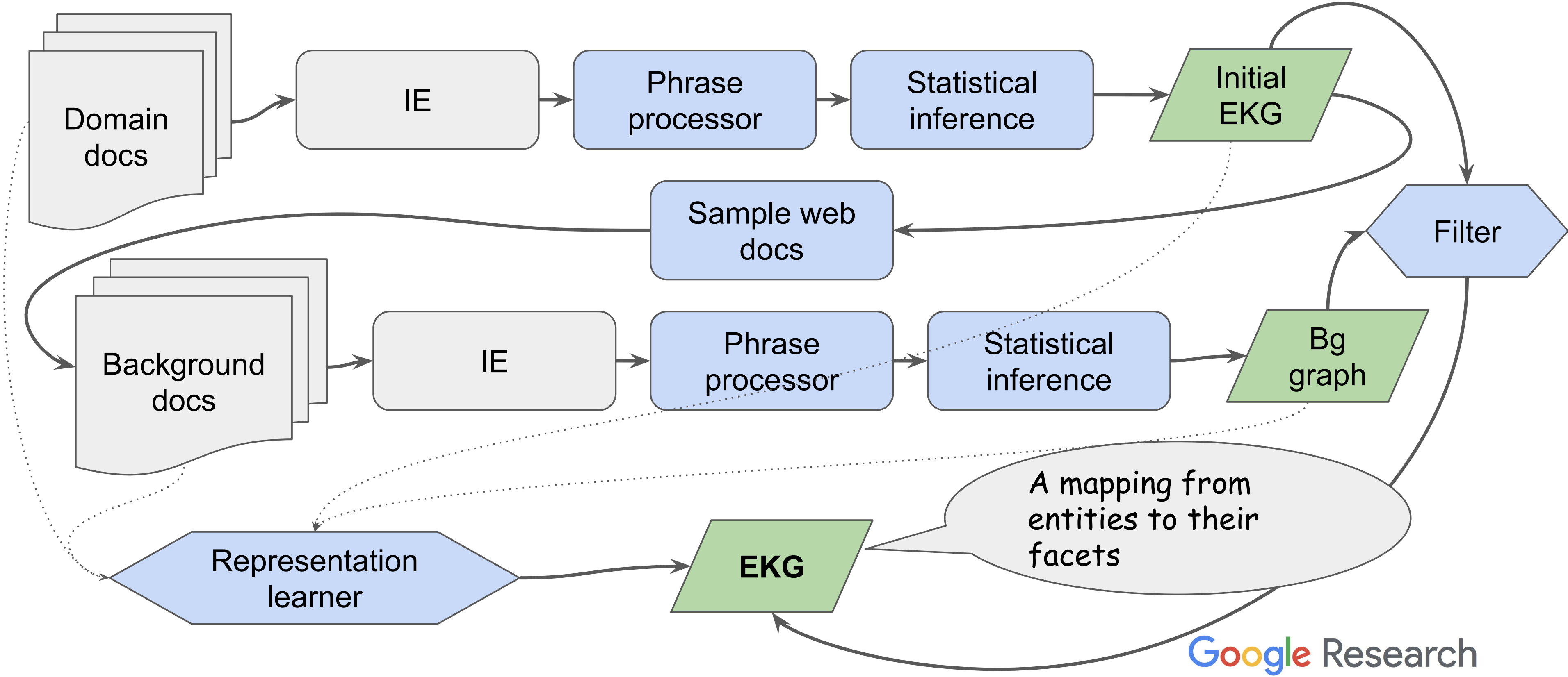
- Etymology
    - *sum, esse*
- Historical NLP-based definition: named entities, coarse-grained types
- Instances
    - Albert Einstein is an instance of a PERSON
    - … but what about “Alpha-1-fetoprotein”?
- “Thing that can appear as a node in a KG”
- Grounding
- Abstraction that
    - … provides succinct access to a bundle of information
    - … humans have succinct referring expressions for

Where is an entity?

Look at this from two points of view: KG induction and multilinguality



EKG induction



Domain docs

IE

Phrase processor

Statistical inference

Initial EKG

Sample web docs

Filter

Background docs

IE

Phrase processor

Statistical inference

Bg graph

Representation learner

EKG

A mapping from entities to their facets

Google Research

EKG induction: closer look

Statistical inference

- uses a generalization of KL divergence, i-divergence (Papineni, 2001)

$$\alpha = \frac{n \cdot c(s,t)}{c(s) \cdot c(t)}$$

$$\mathrm{idiv}(s,t) = c(s,t)\left[\frac{1}{\alpha} - 1.0 + \log(\alpha)\right]$$

Filter

- For each topic in foreground that is also in background
  - retain only topics whose facet dis’ns above JS divergence threshold
  - for each retained foreground topic
    - filter facets based on i-divergence from background

$$\text{Let } a = \frac{p+q}{2}.$$

$$D_{\mathrm{JS}}\left(p \| q\right) = \tfrac{1}{2} D_{\mathrm{KL}}\left(p \| a\right) + \tfrac{1}{2} D_{\mathrm{KL}}\left(q \| a\right).$$

KG induction in practice

Graph building for your corpus

- Cloud AI Workshop Generating Specialized Knowledge Graphs

Domain-specific entities: biomedical text

- COVID-19 Research Explorer

Way back in history...

It’s 2018

[Learning Dense Representations for Entity Retrieval]

Research questions

- Can we learn effective neural mention and entity encoders?
- Can entity linking be formulated as a nearest neighbor retrieval problem?
- Can we learn a high-performing entity linking model with no alias tables?

Joint work of:
Daniel Gillick, Sayali Kulkarni, Larry Lansing,
Alessandro Presta, Jason Baldridge, Eugene Ie, Diego Garcia-Olano

Entity Linking

What is **George Harrison**’s favorite Nintendo game?

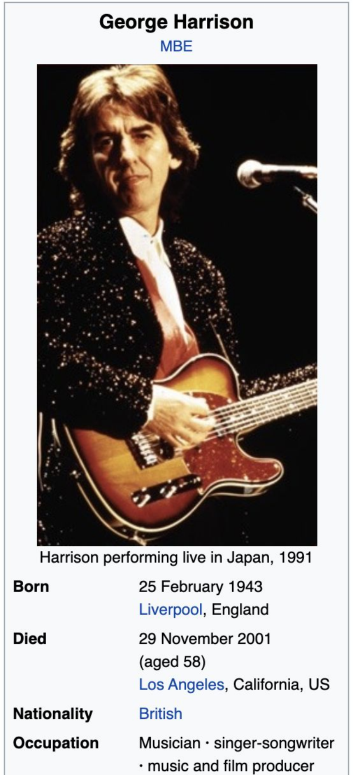
Entity Linking

What is **George Harrison**'s favorite Nintendo game?

~6M English entities

Entity Linking

What is **George Harrison**'s favorite Nintendo game?



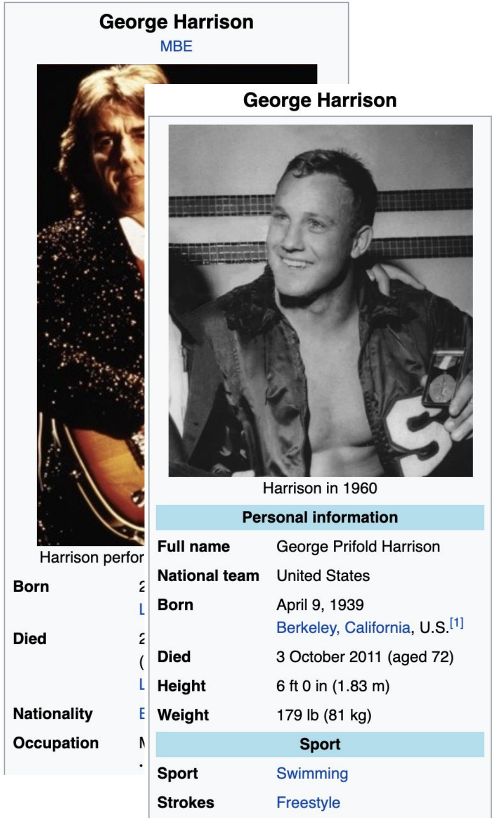
George Harrison
MBE

Harrison performing live in Japan, 1991

Born	25 February 1943 Liverpool, England
Died	29 November 2001 (aged 58) Los Angeles, California, US
Nationality	British
Occupation	Musician · singer-songwriter · music and film producer

Entity Linking

What is **George Harrison**'s favorite Nintendo game?



George Harrison

MBE

Harrison perfor

Born

Died

Nationality

Occupation

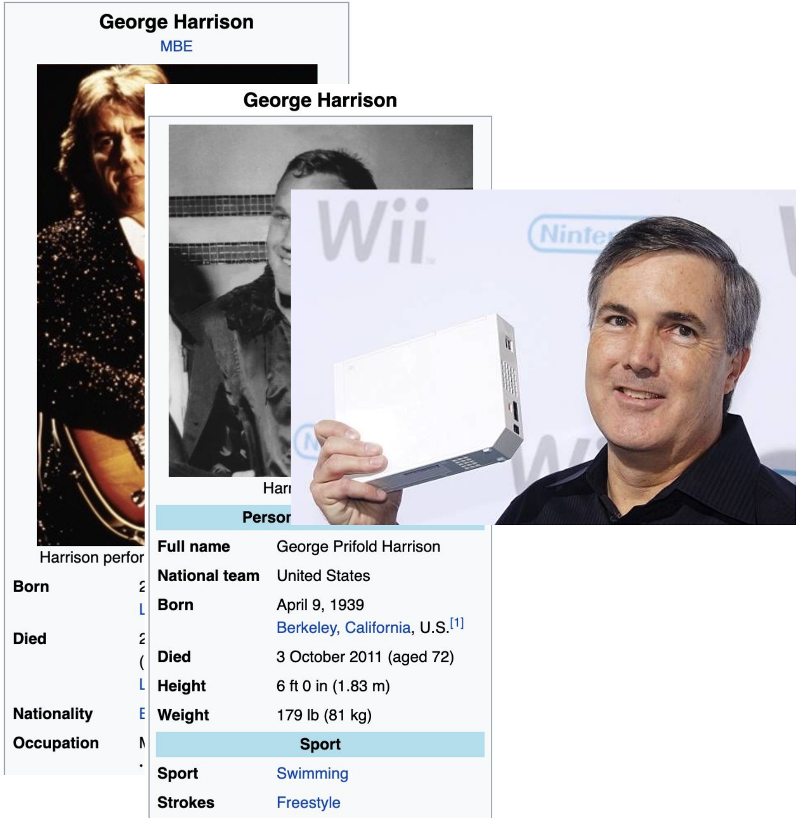
George Harrison

Harrison in 1960

Personal information	
Full name	George Prifold Harrison
National team	United States
Born	April 9, 1939 Berkeley, California, U.S.[1]
Died	3 October 2011 (aged 72)
Height	6 ft 0 in (1.83 m)
Weight	179 lb (81 kg)
Sport	
Sport	Swimming
Strokes	Freestyle

Entity Linking

What is **George Harrison**'s favorite Nintendo game?



George Harrison
MBE

Harrison perfor

Born
Died
Nationality
Occupation

George Harrison

Harr

Persor

Full name	George Prifold Harrison
National team	United States
Born	April 9, 1939 Berkeley, California, U.S.[1]
Died	3 October 2011 (aged 72)
Height	6 ft 0 in (1.83 m)
Weight	179 lb (81 kg)
Sport	
Sport	Swimming
Strokes	Freestyle

Entity Linking

What is **George Harrison**'s favorite Nintendo game?

Arts and entertainment [edit]

- George Henry Harrison (1816–1846), English watercolour painter
- George Harrison Marks (1926–1997), English photographer and dire
- *George Harrison* (album), a 1979 album by George Harrison

Business [edit]

- George L. Harrison (1887–1958), American banker
- George Charter Harrison (1881–1959), Anglo-American managemer
- George Harrison (executive) (fl. c. 2000), American business manag

Politics [edit]

- George Harrison (Hertford MP) (1680–1759), British MP for Hertford
- George Harrison (Bossiney MP), Member of Parliament (MP) for Bos
- George Harrison (civil servant) (1767–1841), British jurist & governm
- George Harrison (Lord Provost) (1811–1885), Lord Provost of Edinb
- George Paul Harrison, Sr. (1813–1888), American politician, Georgi
- George Paul Harrison, Jr. (1841–1922), American politician, U.S. Re
- George Moffett Harrison (1847–1923), American politician in Virginia
- George Harrison (Irish republican) (1915–2004), member of the Prov

Sports [edit]

- George Harrison (Yorkshire cricketer) (1862–1940), British cricketer
- George Harrison (Glamorgan cricketer) (1895–?), English cricketer
- George Harrison (footballer) (1892–1939), professional footballer for
- George Harrison (swimmer) (1939–2011), American swimmer

Other [edit]

- George H. Harrison (1841–1919), American sailor and Medal of Hon
- George R. Harrison (1898–1979), American physicist
- George Harrison (prospector), Australian discoverer of gold in the Tr

See also [edit]

- Harrison George, American activist

Standard approach

Mention + Context → Entity Candidates → Ranked Entities

IR via Alias Table

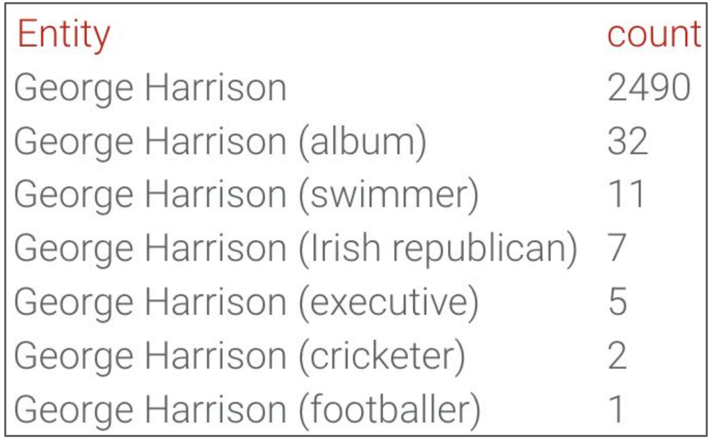
Scoring Model

Standard approach

Mention + Context → Entity Candidates → Ranked Entities

IR via Alias Table

Scoring Model



Entity	count
George Harrison	2490
George Harrison (album)	32
George Harrison (swimmer)	11
George Harrison (Irish republican)	7
George Harrison (executive)	5
George Harrison (cricketer)	2
George Harrison (footballer)	1

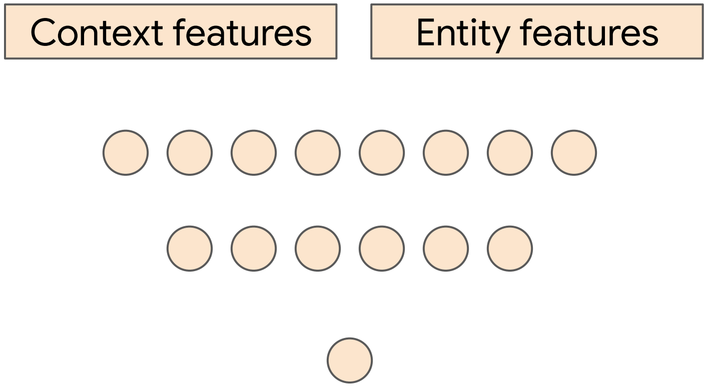
Standard approach

Mention + Context → Entity Candidates → Ranked Entities

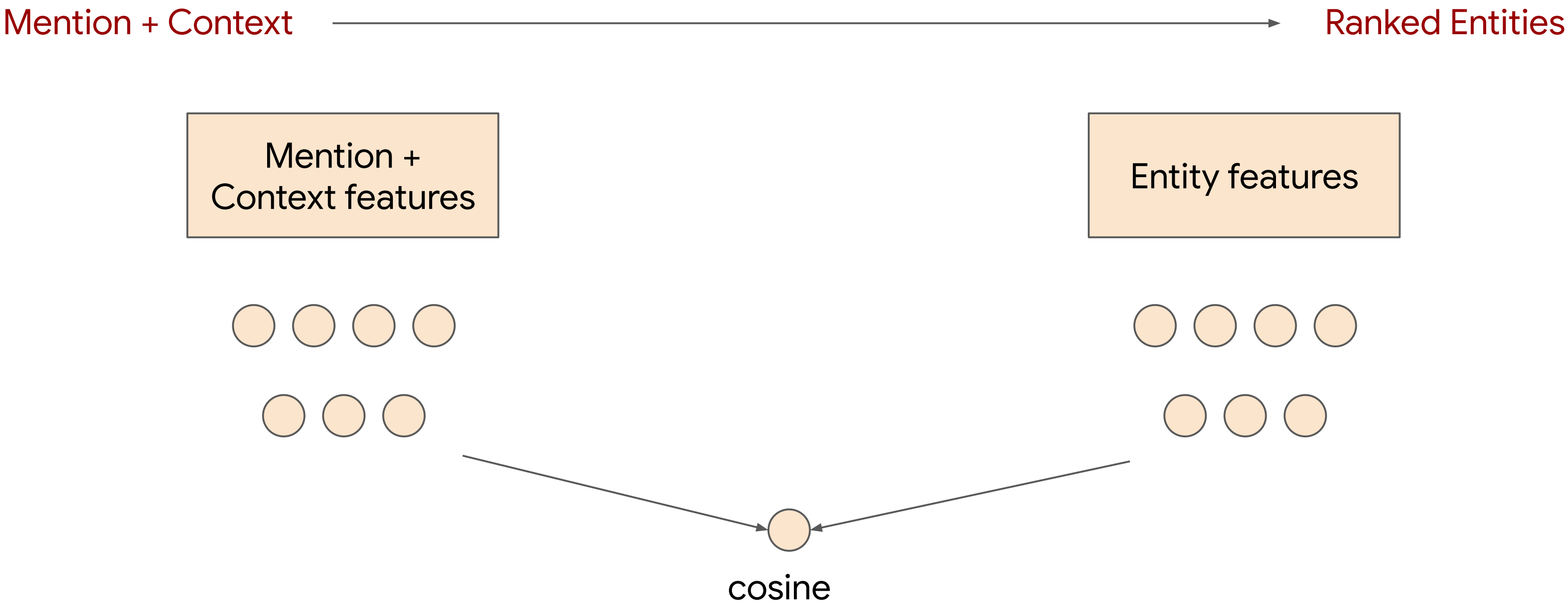
IR via Alias Table

Scoring Model

Entity	count
George Harrison	2490
George Harrison (album)	32
George Harrison (swimmer)	11
George Harrison (Irish republican)	7
George Harrison (executive)	5
George Harrison (cricketer)	2
George Harrison (footballer)	1



End-to-end learned approach



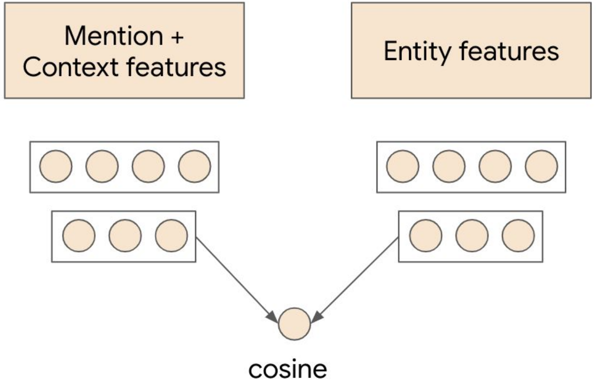
Mention + Context → Ranked Entities

Mention + Context features

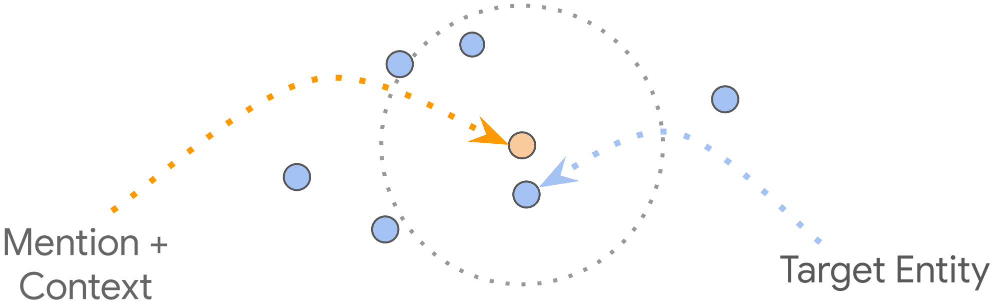
Entity features

cosine

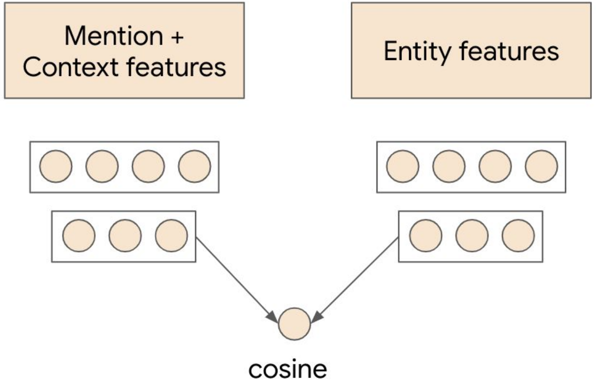
Dual encoder inference



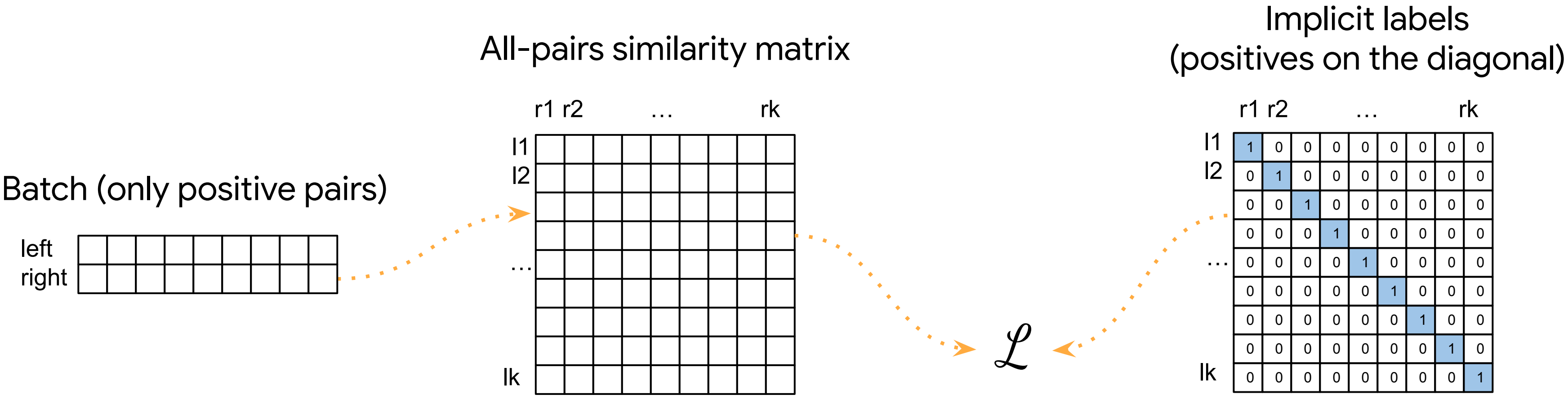
Approximate nearest neighbor search



Dual encoder training



In-batch sampled softmax loss

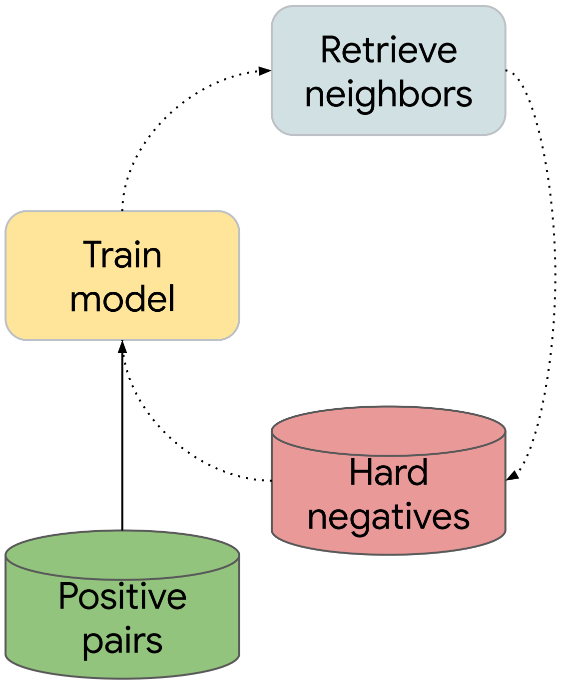


	r1	r2			…				rk
l1	1	0	0	0	0	0	0	0	0
l2	0	1	0	0	0	0	0	0	0
	0	0	1	0	0	0	0	0	0
	0	0	0	1	0	0	0	0	0
…	0	0	0	0	1	0	0	0	0
	0	0	0	0	0	1	0	0	0
	0	0	0	0	0	0	1	0	0
	0	0	0	0	0	0	0	1	0
lk	0	0	0	0	0	0	0	0	1

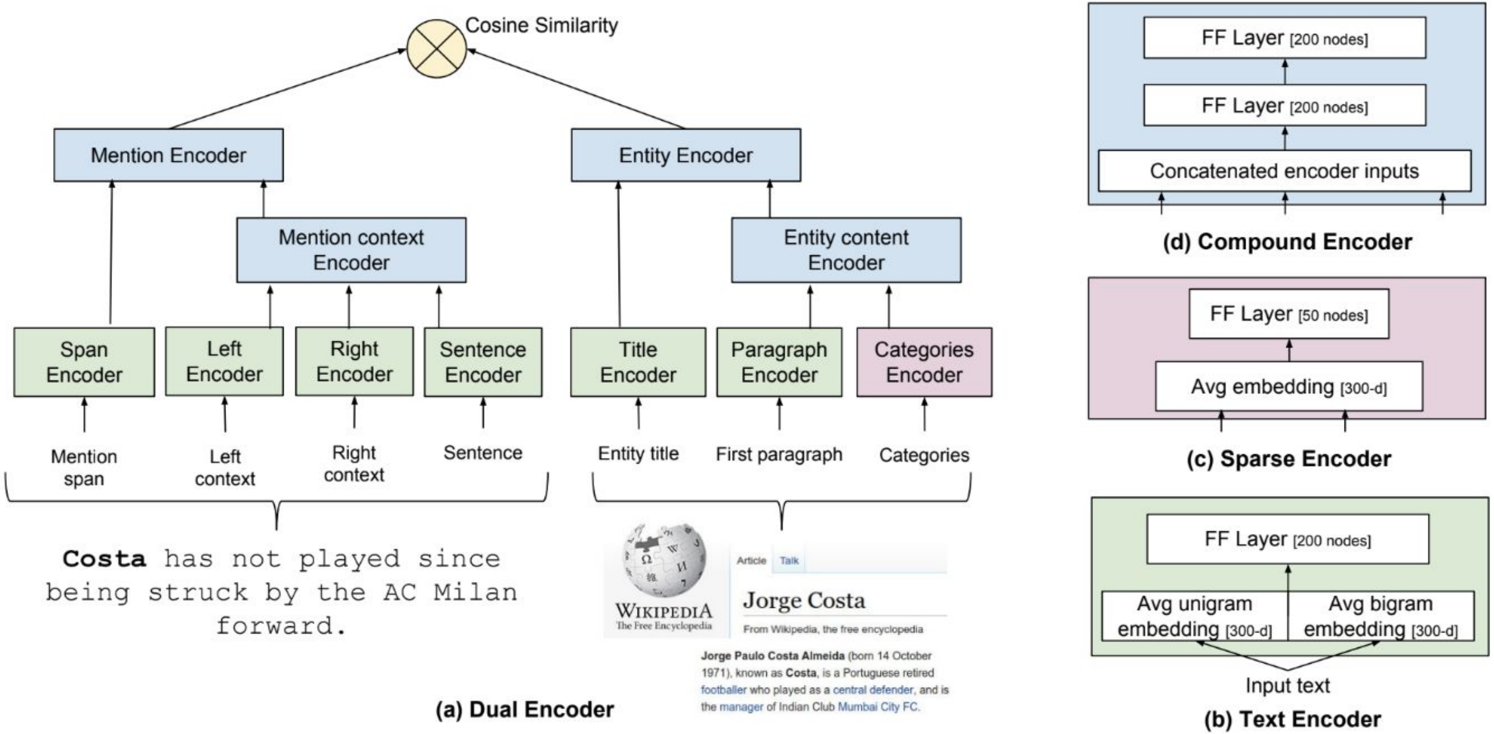
DE training: negative mining

Repeat:

1. Train model
2. Mine negatives
    a. Retrieve nearest neighbors
    b. Select likely negatives



DEER: Dual Encoder for Entity Retrieval and linking



Costa has not played since being struck by the AC Milan forward.

Jorge Costa

From Wikipedia, the free encyclopedia

Jorge Paulo Costa Almeida (born 14 October 1971), known as **Costa**, is a Portuguese retired footballer who played as a central defender, and is the manager of Indian Club Mumbai City FC.

(a) Dual Encoder

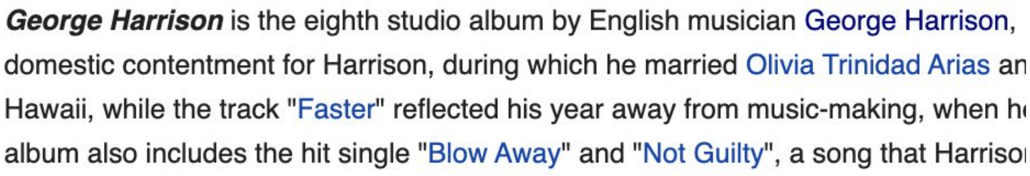
(d) Compound Encoder

(c) Sparse Encoder

(b) Text Encoder

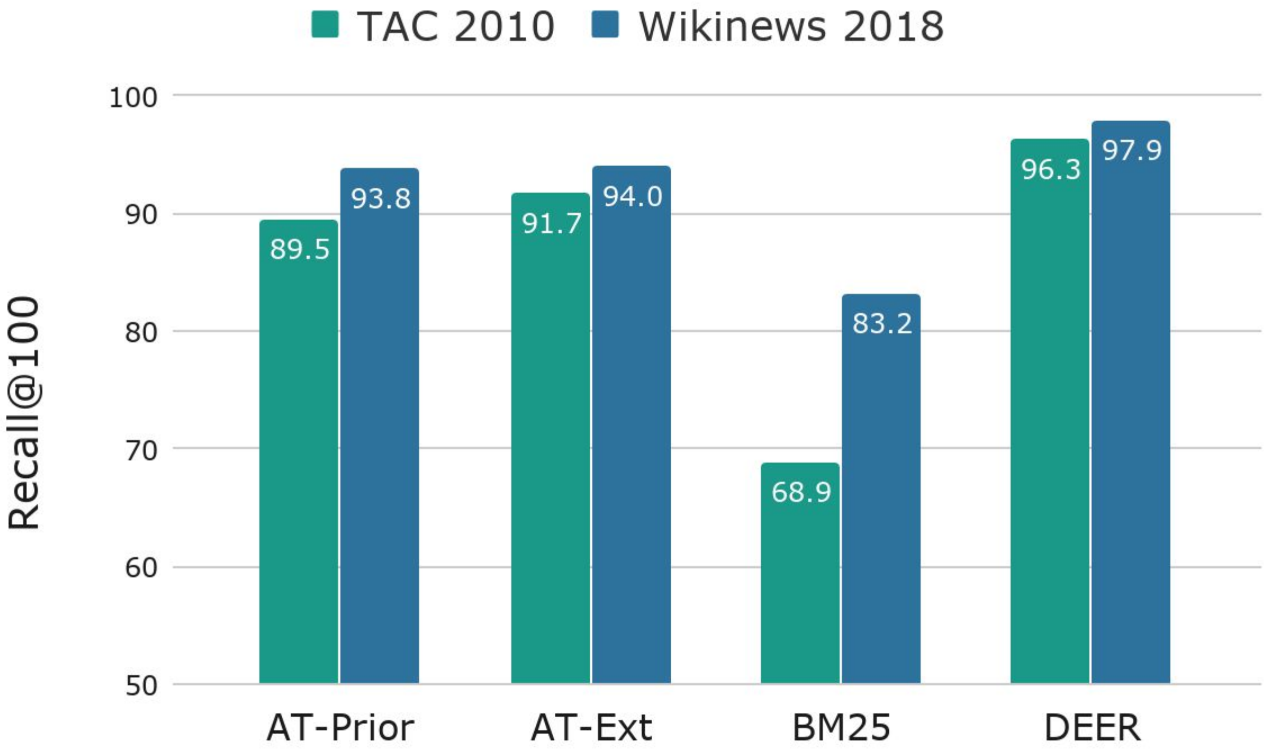
Data

- Training
  - Wikipedia hyperlinks (10/2018)
    - 5.7M entities
    - 113M linked mentions
- Evaluation
  - TAC 2010 corpus
    - 1024 linked mentions
  - Wikinews 2018
    - 2263 linked mentions
    - [Open-sourced]()
  - Gold mentions provided
  - No NIL entities
  - No in-domain tuning



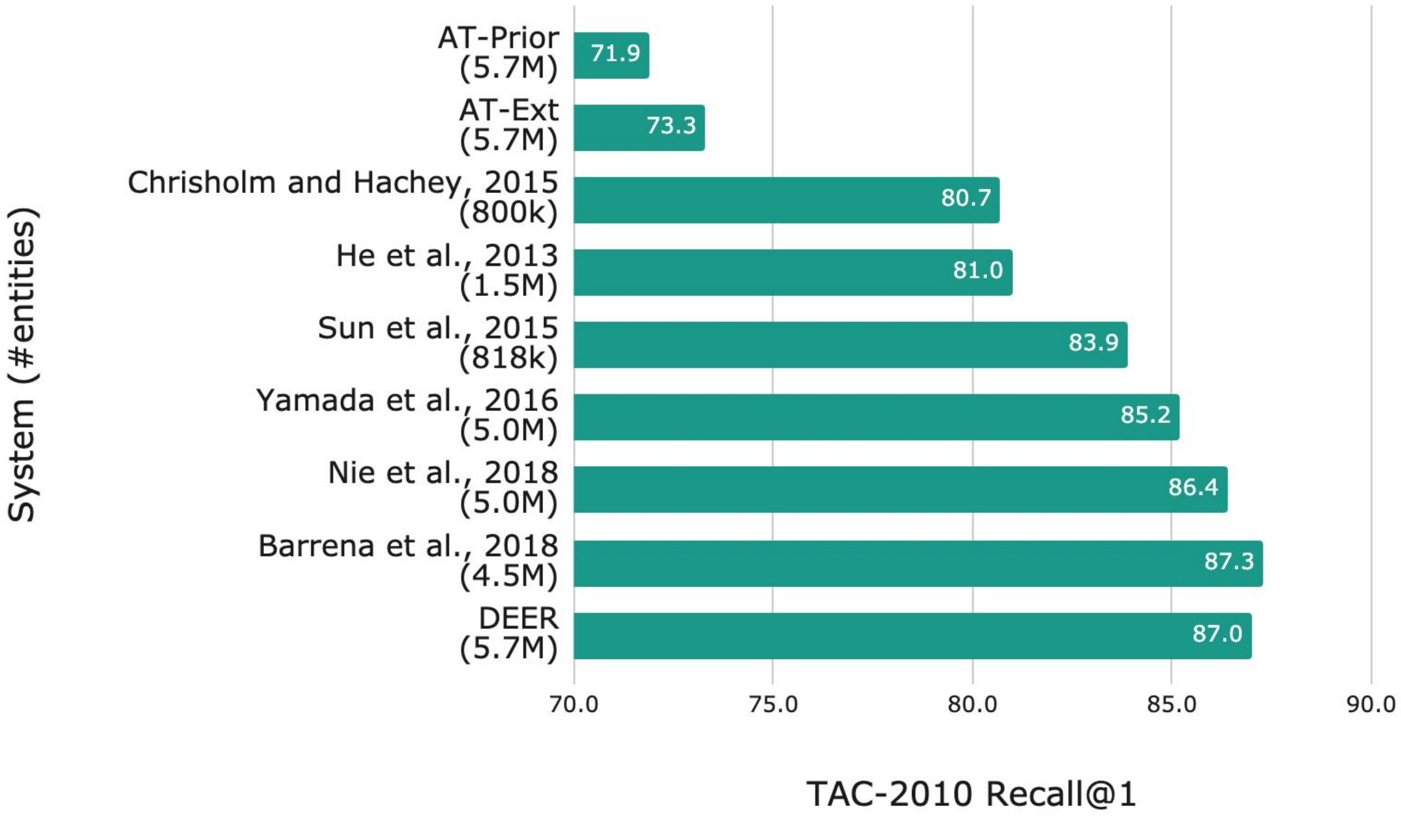
George Harrison is the eighth studio album by English musician George Harrison, domestic contentment for Harrison, during which he married Olivia Trinidad Arias an Hawaii, while the track "Faster" reflected his year away from music-making, when h album also includes the hit single "Blow Away" and "Not Guilty", a song that Harriso

Results

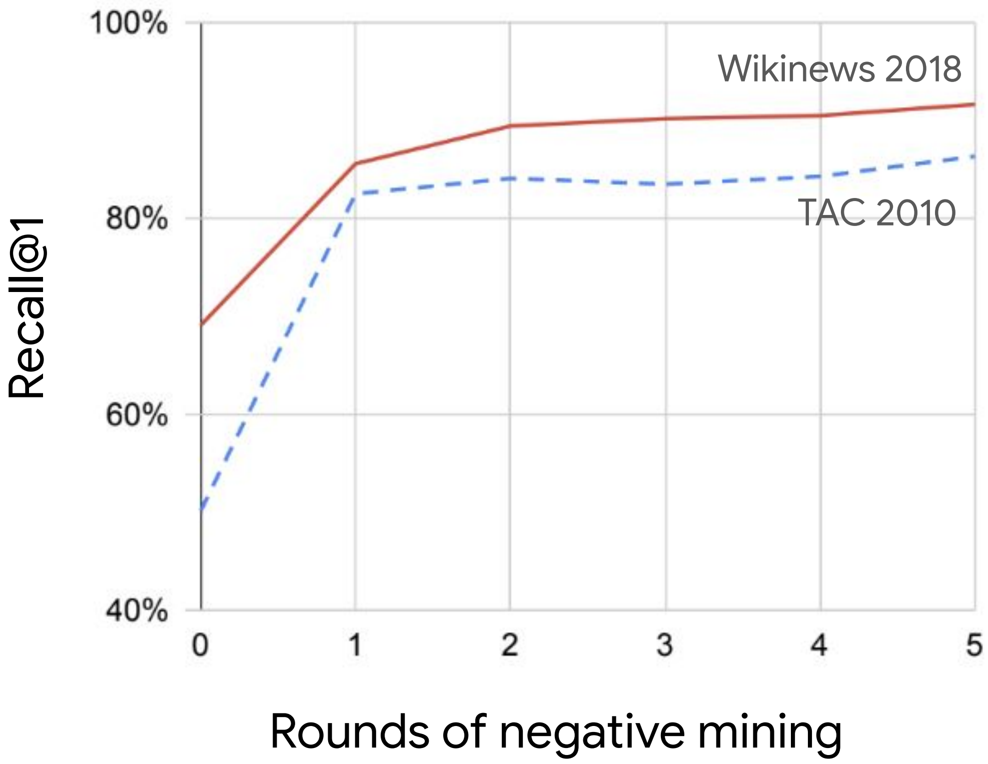


Google Research

Results


System (#entities)
AT-Prior (5.7M)
71.9
AT-Ext (5.7M)
73.3
Chrisholm and Hachey, 2015 (800k)
80.7
He et al., 2013 (1.5M)
81.0
Sun et al., 2015 (818k)
83.9
Yamada et al., 2016 (5.0M)
85.2
Nie et al., 2018 (5.0M)
86.4
Barrena et al., 2018 (4.5M)
87.3
DEER (5.7M)
87.0
70.0
75.0
80.0
85.0
90.0
TAC-2010 Recall@1

Results



Recall@1
100%
80%
60%
40%
Wikinews 2018
TAC 2010
0
1
2
3
4
5
Rounds of negative mining

From retrieval to ranking

If you can formulate an alias table–free way of retrieving entities, then

- why not use a cross-attention model to rerank?

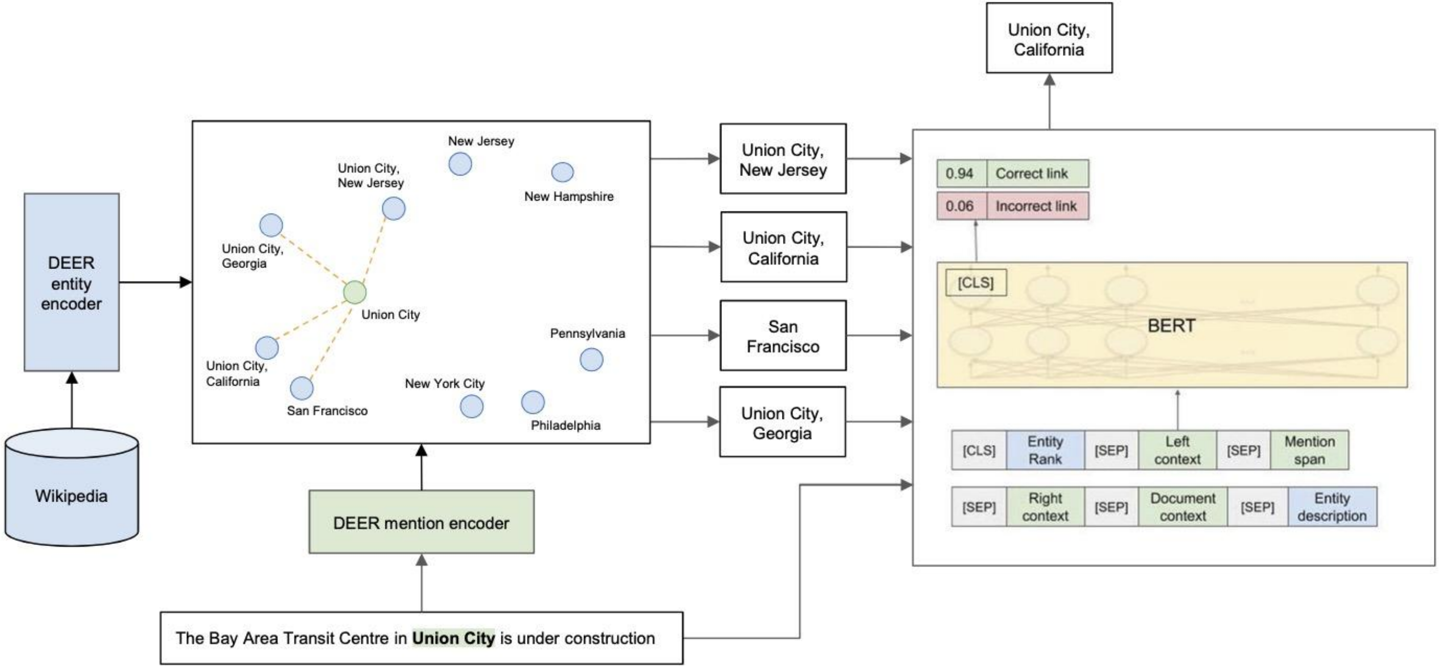
Joint work with Oshin Agarwal

[Entity Linking via Dual and Cross-Attention Encoders]()

Reranking candidates

- BERT classifier trained on domain data
- Training: 10 nearest neighbors and true link
- Inference: 100 nearest neighbors from DEER
- Predicted entity: candidate with highest class probability

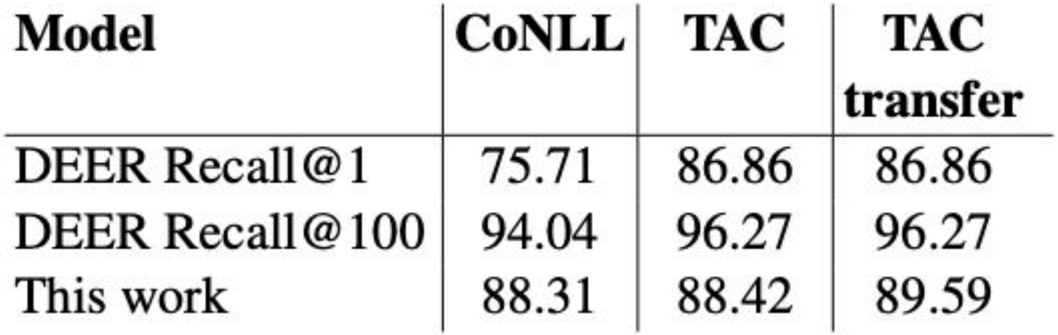
- Document context
    - ○ Other mentions
    - ○ Bag of Words



The Bay Area Transit Centre in **Union City** is under construction

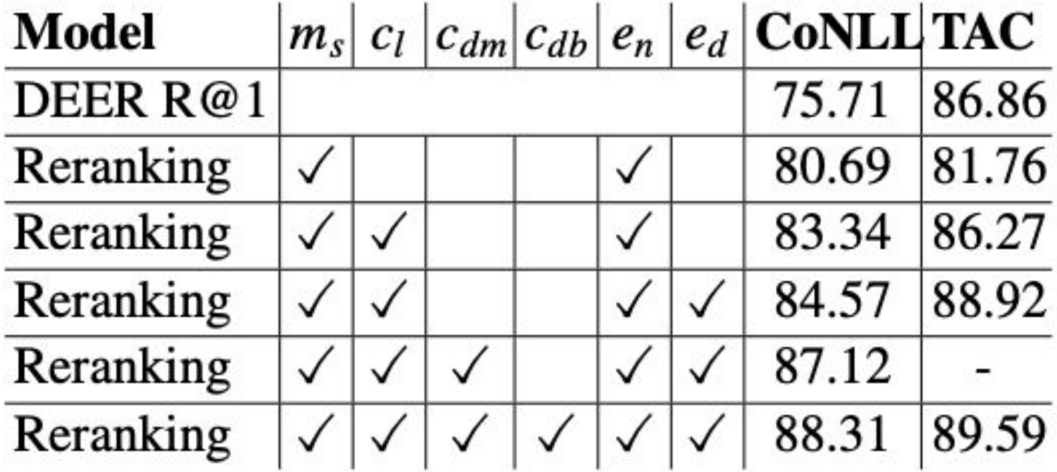
Results

- Accuracy (Recall@1) on CoNLL ’03 and TAC KBP ’10
    - Both the datasets had updated Wikipedia links according to the Wikipedia dump used



Model	**CoNLL**	**TAC**	**TAC transfer**
DEER Recall@1	75.71	86.86	86.86
DEER Recall@100	94.04	96.27	96.27
This work	88.31	88.42	89.59

Ablation



Model	m_s	c_l	c_{dm}	c_{db}	e_n	e_d	**CoNLL**	**TAC**
DEER R@1							75.71	86.86
Reranking	✓				✓		80.69	81.76
Reranking	✓	✓			✓		83.34	86.27
Reranking	✓	✓			✓	✓	84.57	88.92
Reranking	✓	✓	✓		✓	✓	87.12	-
Reranking	✓	✓	✓	✓	✓	✓	88.31	89.59

m_s is mention surface form

c_l is local context

c_{dm} is other mentions in doc

c_{db} is doc bag-of-words

e_n is entity name

e_d is entity description

But there are far more entities than these

If we continue to use the rapidly aging entity linking academic datasets, we miss too much

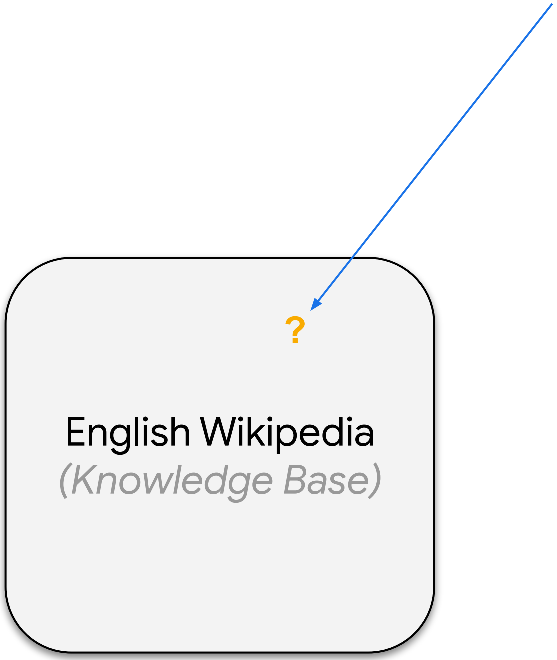
Let's do truly multilingual entity linking in 100 languages

Jan Botha
jabot@
google.com

Zifei Shan
zifeishan@
gmail.com

Daniel Gillick
dgillick@
google.com

“The torii gate next to ***Tsuru no Yu Onsen*** *is the starting point for this hike.”*



?

English Wikipedia

(Knowledge Base)

“The torii gate next to ***Tsuru no Yu Onsen*** *is the starting point for this hike.”*

Only Japanese Wikipedia has this entity!

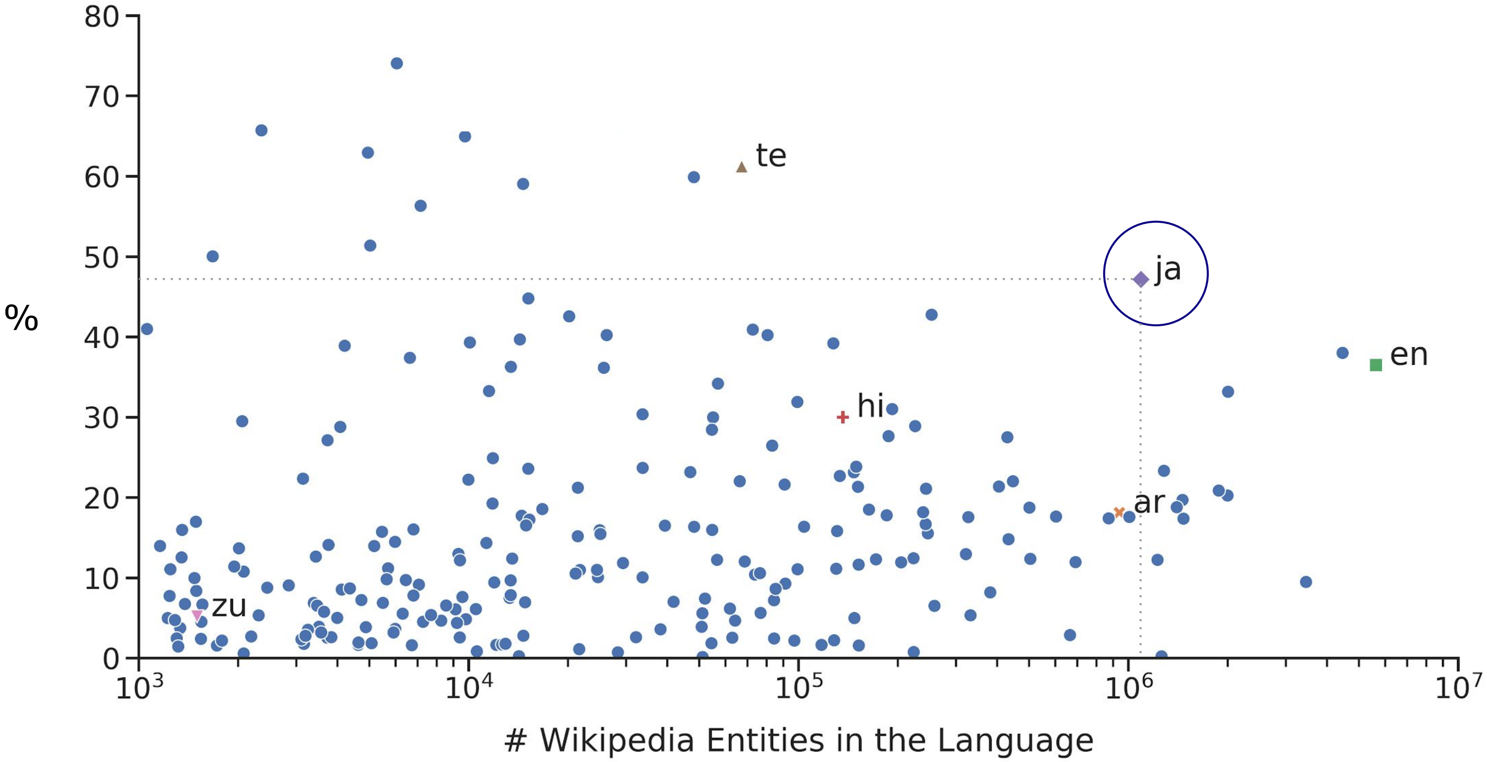
鶴の湯温泉は、秋田県仙北市の乳頭温泉郷内にある温泉である。

Translation: Tsuru no Yu Onsen is a hot spring located in the Tae no Yu hot spring village in Semboku City, Akita Prefecture.

English Wikipedia
(Knowledge Base)

Source: Wikimedia Commons

Percentage of Language-unique Wikipedia entities



Excludes redirects and special pages (lists, categories, disambiguation etc). Showing 231 languages with more than 1,000 entities as of October 2019 (WikiData & Wikipedia)

Entity Linking Variants

Monolingual

[en] Murray wrote the music and **Callander** the lyrics.

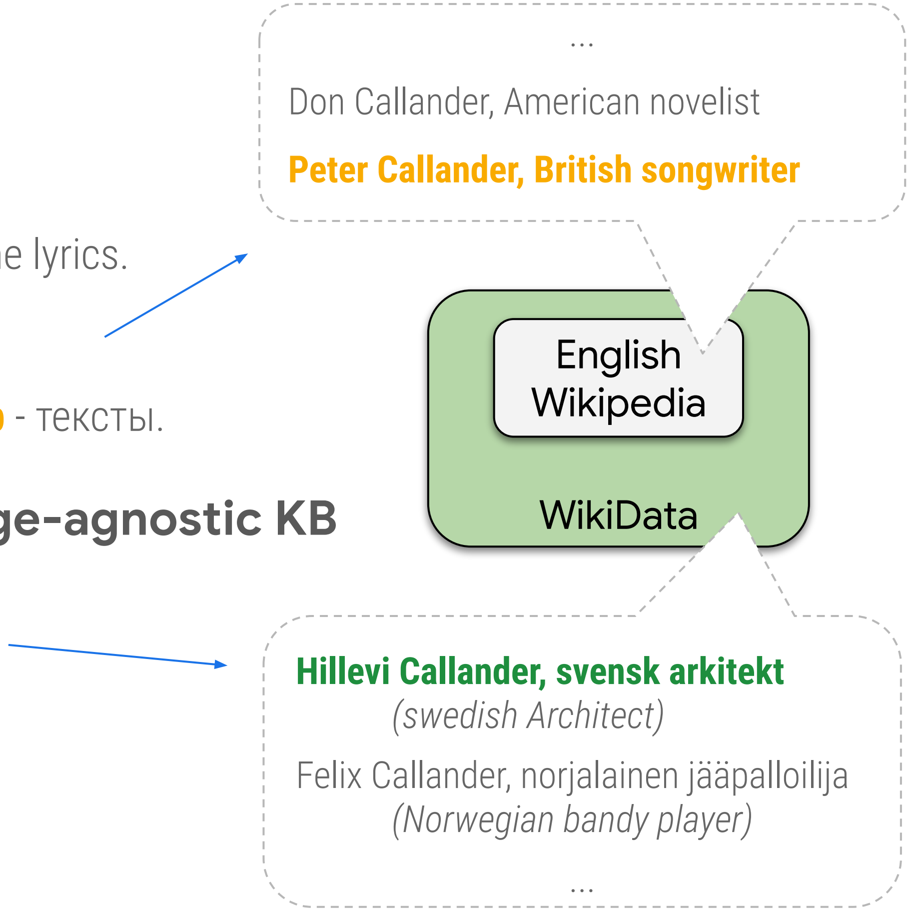
Cross-Lingual (XEL)

[ru]: Мюррей написал музыку, а **Калландер** - тексты.

Multilingual Entity Linking to language-agnostic KB

[en]: **Callander** studied in Stockholm.

[ru]: **Калландер** учился в Стокгольме.



...

Don Callander, American novelist

Peter Callander, British songwriter

English Wikipedia

WikiData

Hillevi Callander, svensk arkitekt
(swedish Architect)

Felix Callander, norjalainen jääpalloilija
(Norwegian bandy player)

...

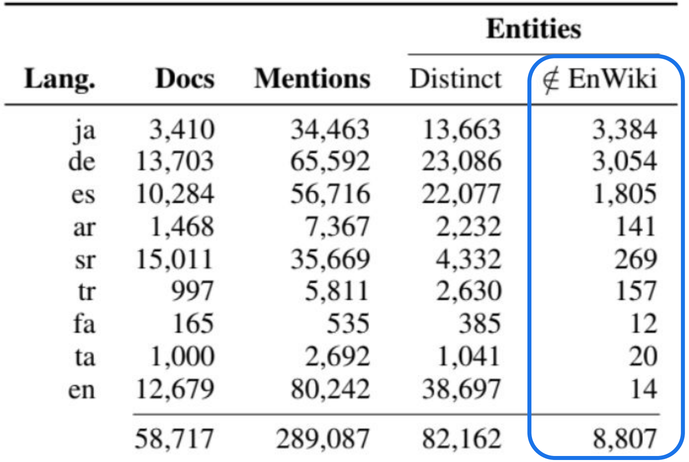
Multilingual Entity Linking to a Language-Agnostic KB

- Systematically take this setting to its logical conclusion
- KB = 20m WikiData entities
- Each entity represented with its Wikipedia description
    - sourced from 104 languages in a data-driven way
- Large-scale supervision: naturally occurring Wikipedia hyperlinks
    - >600m of these

Mewsli-9: *Multilingual Entities in News, linked*

- New evaluation set for Multilingual EL
- 290k mentions ~ 82k WikiData entities
- **Many entities outside English Wikipedia**

  → success *requires* using expanded KB



			Entities	
Lang.	**Docs**	**Mentions**	Distinct	∉ EnWiki
ja	3,410	34,463	13,663	3,384
de	13,703	65,592	23,086	3,054
es	10,284	56,716	22,077	1,805
ar	1,468	7,367	2,232	141
sr	15,011	35,669	4,332	269
tr	997	5,811	2,630	157
fa	165	535	385	12
ta	1,000	2,692	1,041	20
en	12,679	80,242	38,697	14
	58,717	289,087	82,162	8,807

Mewsli-9: *Multilingual Entities in News, linked*

- New evaluation set for Multilingual EL
- 290k mentions ~ 82k WikiData entities
- **Many entities outside English Wikipedia**

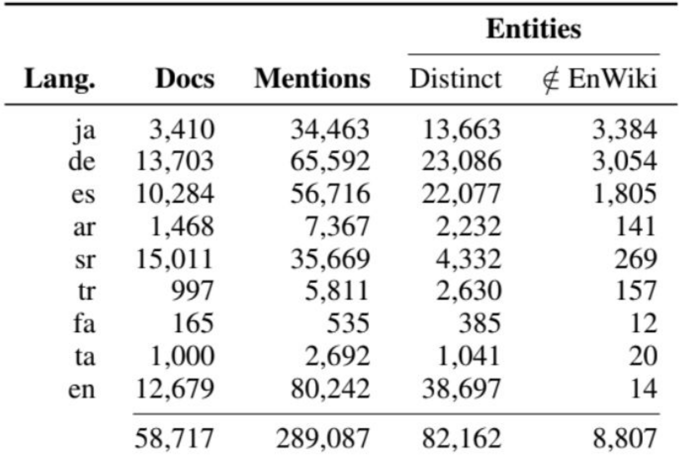
  → success *requires* using expanded KB
- Extracted from WikiNews.org articles
- Linguistic diversity:

  9 languages

  5 language families

  6 orthographies

goo.gle/mewsli-dataset

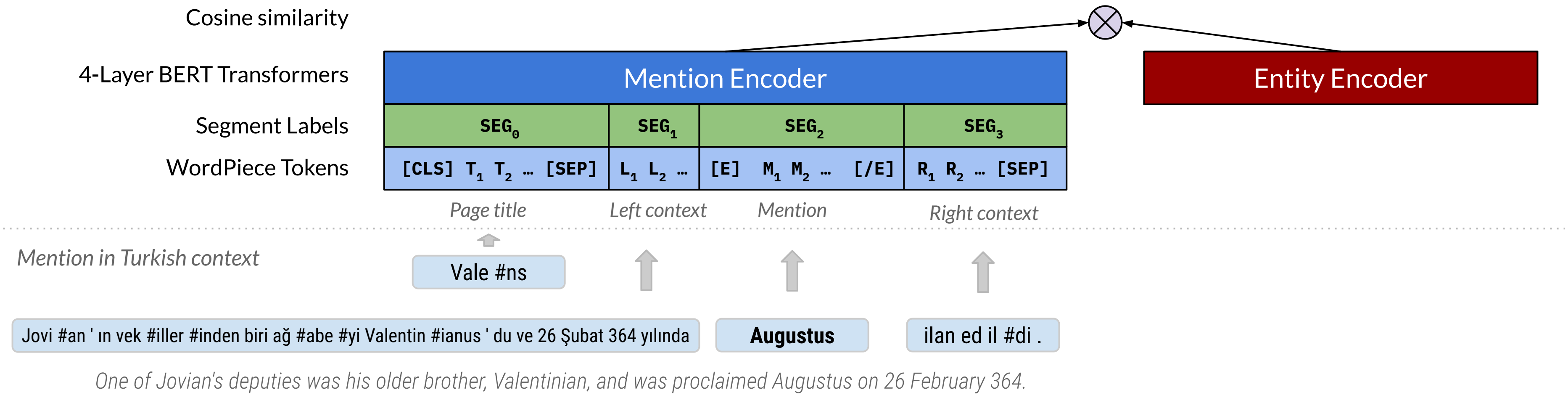


			Entities	
Lang.	**Docs**	**Mentions**	Distinct	∉ EnWiki
ja	3,410	34,463	13,663	3,384
de	13,703	65,592	23,086	3,054
es	10,284	56,716	22,077	1,805
ar	1,468	7,367	2,232	141
sr	15,011	35,669	4,332	269
tr	997	5,811	2,630	157
fa	165	535	385	12
ta	1,000	2,692	1,041	20
en	12,679	80,242	38,697	14
	58,717	289,087	82,162	8,807

Approach & Selected Results

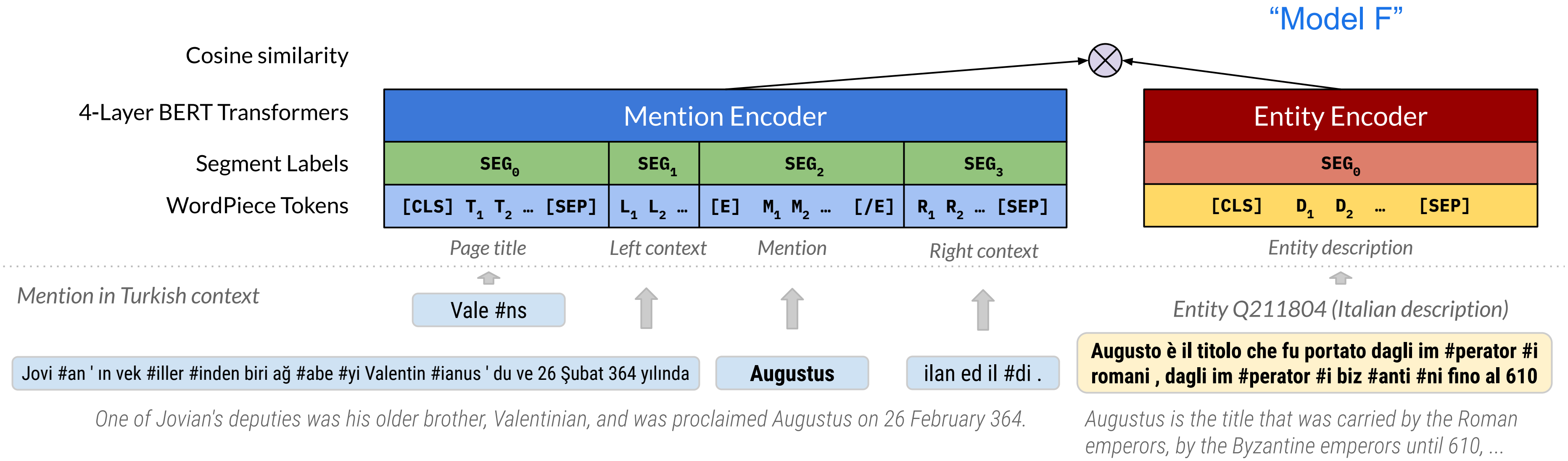
Dual Encoder Model

- **One-step linking** over 20m entities; no IR-style candidate generation



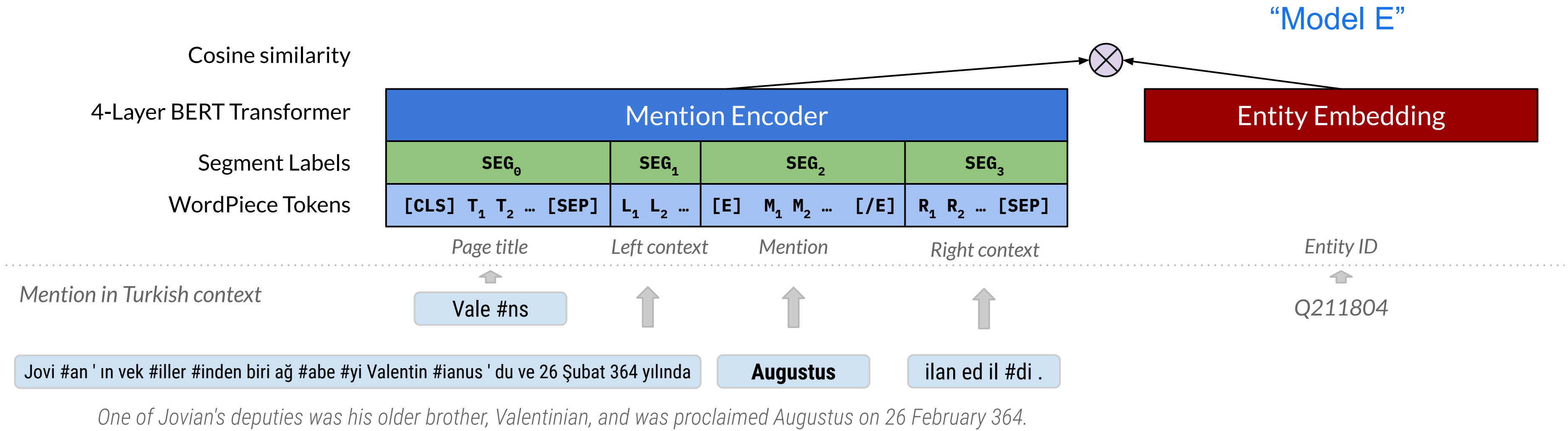
Dual Encoder Model

- **One-step linking** over 20m entities; no IR-style candidate generation
- Model F: Entity featurized with an informative **text description**

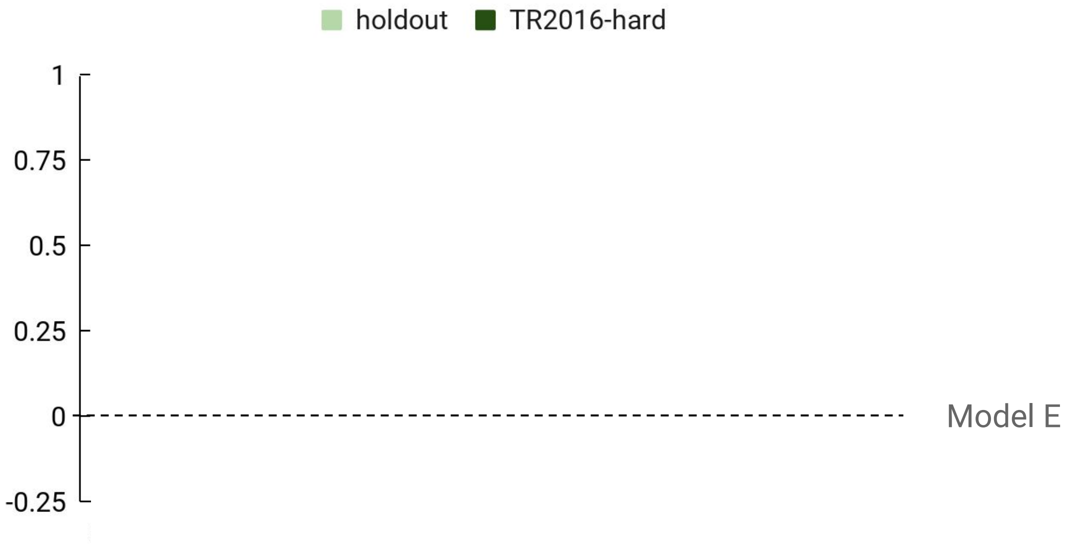


Dual Encoder Model

- **One-step linking** over 20m entities; no IR-style candidate generation
- Model F: Entity featurized with an informative **text description**
- Model E: Entity embedding (baseline)

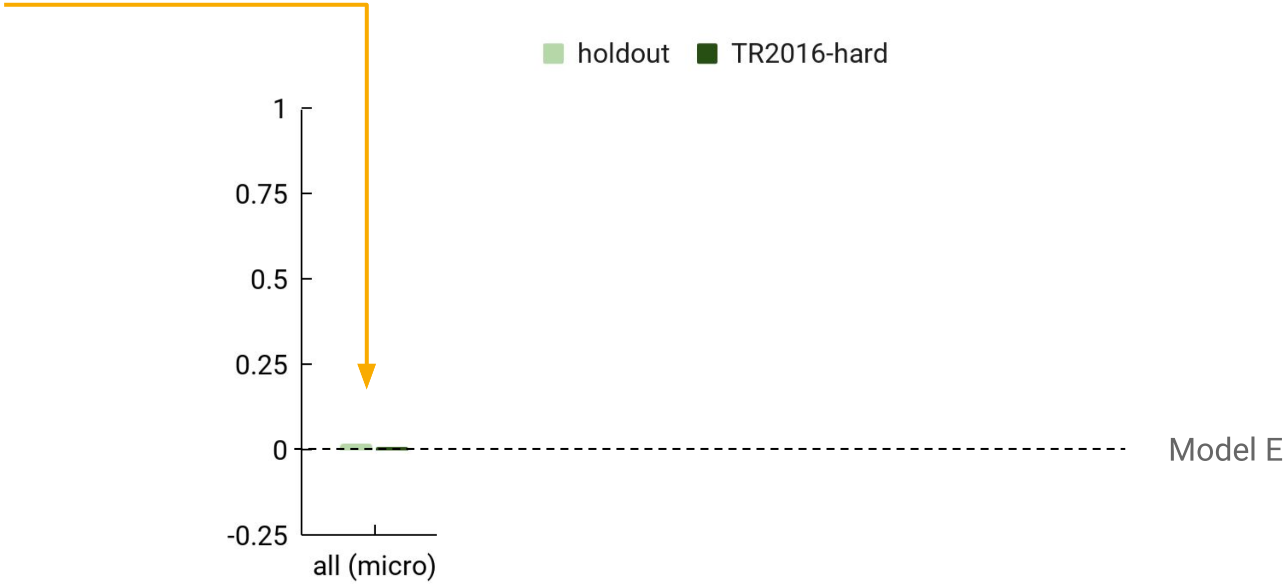


A) Recall@100 increase/decrease for Model F w.r.t. Model E



A) Recall@100 increase/decrease for Model F w.r.t. Model E

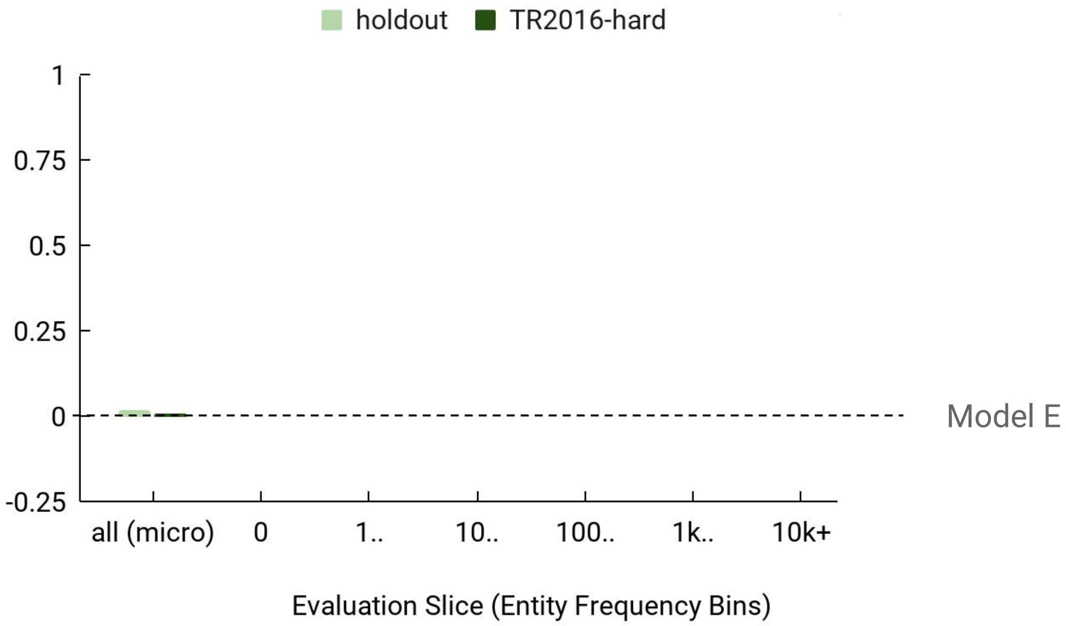
Indistinguishable?



A) Recall@100 increase/decrease for Model F w.r.t. Model E

Indistinguishable?

Evaluate more closely!

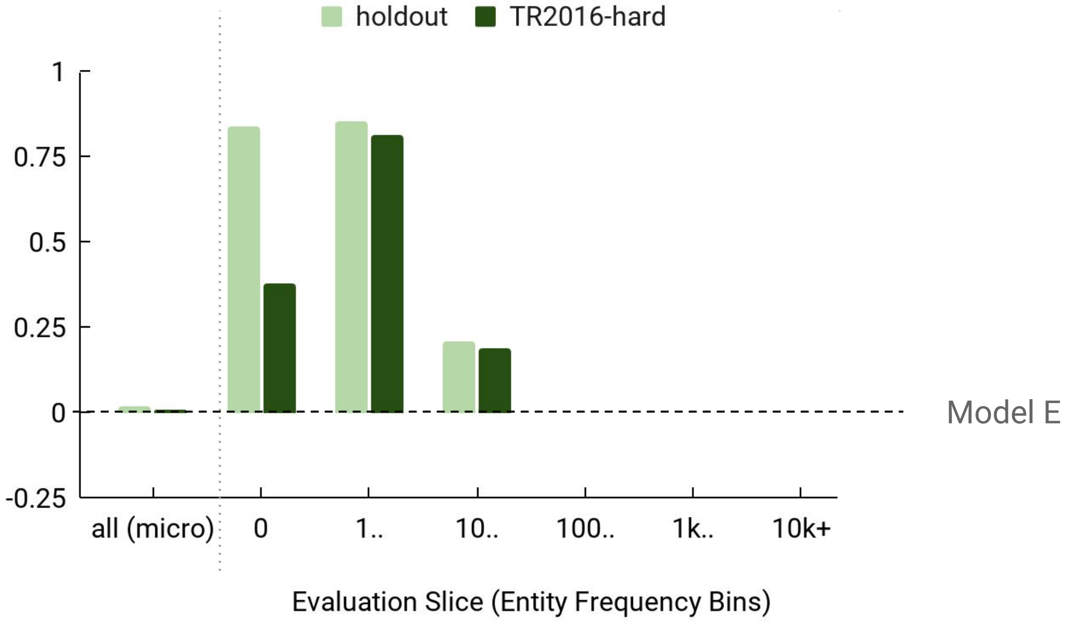


A) Recall@100 increase/decrease for Model F w.r.t. Model E

Indistinguishable?

Evaluate more closely!

- large improvements on unseen & rare entities



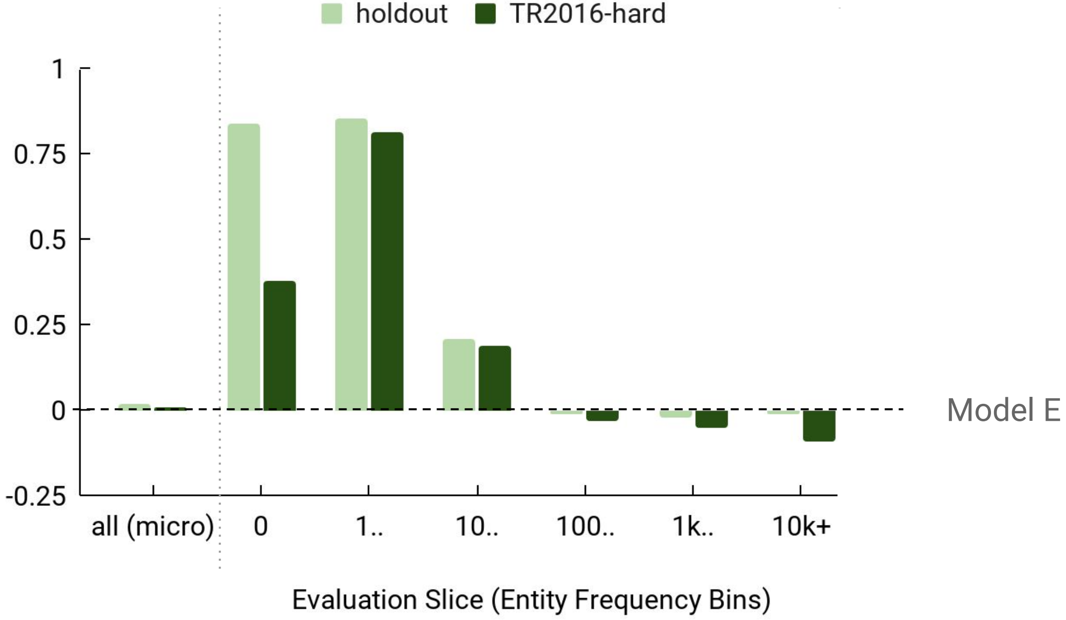
A) Recall@100 increase/decrease for Model F w.r.t. Model E

Indistinguishable?

Evaluate more closely!

- \+ large improvements on unseen & rare entities
- \- negligible decrease on more common entities

Compact generalization



B) Model F+ vs. previous XEL models (on TR2016-hard eval set)

much more limited setting

		Tsai & Roth (2016)	**Upadhyay et al. (2018)**	**Our Model F+**
Setting	Languages	13	5	104
	Entity Vocabulary	5m	5m	20m
	Inference Candidates	20	20	20m
Accuracy	de	0.53	0.55	**0.62**
	es	0.54	0.57	**0.58**
	fr	0.48	0.51	**0.54**
	it	0.48	0.52	**0.56**
	Average	0.51	0.54	**0.57**

B) Model F+ vs. previous XEL models (on TR2016-hard eval set)

		Tsai & Roth (2016)	Upadhyay et al. (2018)	Our Model F+
Setting	Languages	13	5	104
	Entity Vocabulary	5m	5m	20m
	Inference Candidates	20	20	20m
Accuracy	de	0.53	0.55	**0.62**
	es	0.54	0.57	**0.58**
	fr	0.48	0.51	**0.54**
	it	0.48	0.52	**0.56**
	Average	0.51	0.54	**0.57**

proposed approach still outperforms state-of-the art

Gold entity not in English Wikipedia

1) Example Prediction on new Mewsli-9 Dataset

Input (German)

[CLS] Neue Bahnen für den Jenaer Nahverkehr [SEP] Wert auf das neue Design und die technische Ausstattung gelegt . Bei den neuen Bahnen handelt es sich um das Model { **Tramino** } von der polnischen Firma Solaris Bus & Coach . Das Model wurde 2009 vorgestellt und hat sich bei der Ausschreibung durch [SEP]

... the Model { Tramino } from the Polish firm...

✓ **Retrieved entity 1 (Q780281; Polish entity description)**

Solaris Tramino -- rodzina tramwajów , które są produkowane przez firmę Solaris Bus & Coach z Bolechowa koło Poznania .

=> effective cross-lingual retrieval

Source: Wikimedia Commons

2) Example Prediction on new Mewsli-9 Dataset

Gold entity not in English Wikipedia

Input (Serbian)

[CLS] Морали смо да победимо , али смо лоше по [SEP] Душан Ивковић рекао је да је његов тим имао императив победе над { **Италијом** } на Европском првенству , али је утакмицу почео лоше . " Рекао [SEP]

...Dušan Ivković said his team had to beat { Italy } at the European Championship...

✗ **Retrieved entity 1 (Q261190; Italian entity description)**

La nazionale di *pallanuoto* maschile dell' Italia è la squadra di pallanuoto che rappresenta l' Italia nelle competizioni internazionali ; è posta sotto la giurisdizione della Federazione Italiana Nuoto .

✗ **Retrieved entity 2 (Q676899; Italian entity description)**

La nazionale di *calcio* dell' Italia è la selezione maggiore maschile di calcio della Federazione Italiana Giuoco Calcio , il cui nome ufficiale è nazionale A , che rappresenta l' Italia nelle varie competizioni ufficiali o amichevoli riservate a squadre nazionali .

✓ **Retrieved entity 3 (Q734750; Italian entity description)**

La nazionale di *pallacanestro* italiana è la selezione dei migliori giocatori di nazionalità italiana , viene gestita dalla FIP e partecipa ai tornei internazionali di pallacanestro per nazioni gestiti dalla FIBA .

Source: Wikimedia Commons

=> plausible confusion for ambiguous, metonymic mention

Recap: multilingual entity linking

- New task formulation

  multilingual entity linking against language-agnostic KB

- One-step linking feasible: 1 model ~ 104 languages ~ 20m entities
- Fine-grained evaluation important to guide development & analysis
- Mewsli-9: Large and diverse new evaluation dataset to spur further research

Knowledge, LM’s and NLU

So far, we have looked at finding entities. We can

- discover entities and their relationships through distributional analysis
- employ truly multilingual KB’s and data

We have also seen how we can

- model entities well in a dual-encoder setting and using cross-attention models
- can employ feature-based representations of entities, to help in few-shot or zero-shot settings

But can we find new ways to *use* entities and similar “knowledge” in the context of large LM’s and downstream NLU systems and tasks?

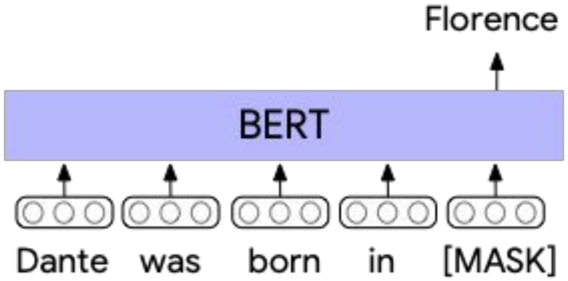
Entities as Experts

What if we build a single model that can employ mention detection, an external entity/knowledge “memory” and large-scale language modeling?

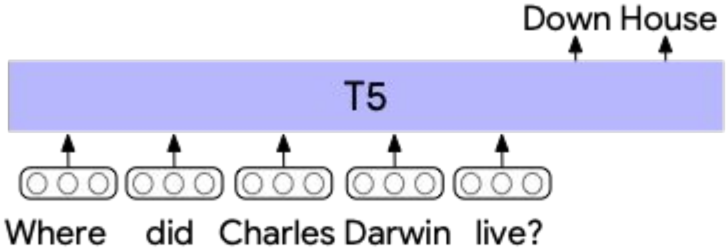
Entities as Experts

Joint work of:
Thibault Févry, Livio Baldini Soares, Nicholas FitzGerald, Eunsol Choi,
Tom Kwiatkowski

Large language models capture world knowledge



Language models as knowledge bases
Petroni et.al. 2019



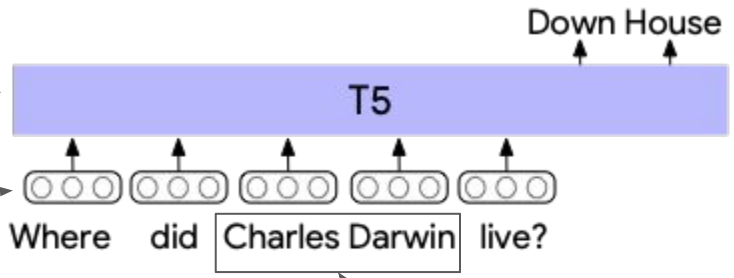
How Much Knowledge Can You Pack Into the Parameters of a Language Model?

Roberts et.al. 2020

Where is the knowledge stored?

11B parameters
Same parameters used for every input.

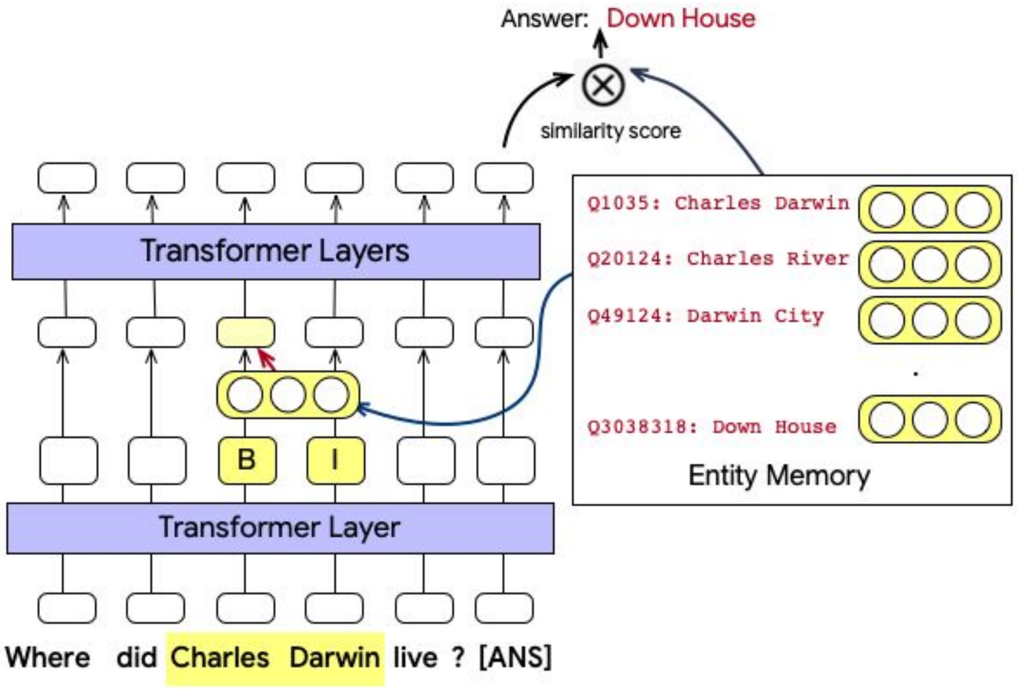
30M parameters
Parameters *conditionally activated* for specific words.



Model has no distinct or independent representation of the key entities.

Entities as Experts (EaE)

- Enhance the transformer with an *entity memory* that contains *distinct and independent* representations of entities.
- Access entity memories *conditionally* — only when needed.
- We hypothesize that this is a *more efficient use of parameters* than a straight-forward transformer stack.



Google Research

Relation to Previous Work

Sparse memory access in sequence models

- Outrageously Large Neural Networks - Shazeer et.al. 2017
- Large Memory Layers with Product Keys - Lample et.al. 2019

EaE adds the extra constraint that memories should be linked to specific entities.

Adding entity representations to sequence models

- Knowledge Enhanced Contextual Word Representations - Peters et.al. 2019
- ERNIE: Enhanced Language Representation with Informative Entities - Zhang et.al. 2019

EaE learns entity memories as part of the sequence model, rather than integrating pre-existing entity representations.

Relation to Previous Work

Sparse memory access in sequence models

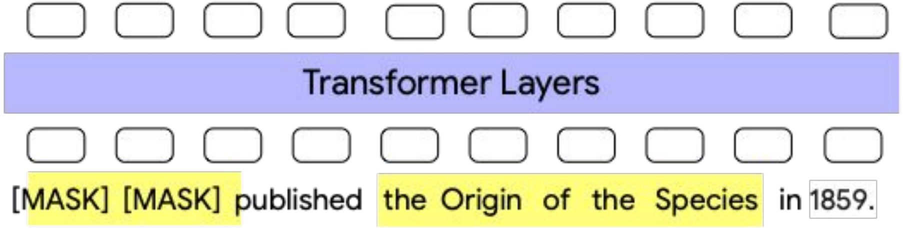
- Outrageously Large Neural Networks - Shazeer et.al. 2017
- Large Memory Layers with Product Keys - Lample et.al. 2019

EaE adds the extra constraint that memories should be linked to specific entities.

Adding entity representations to sequence models

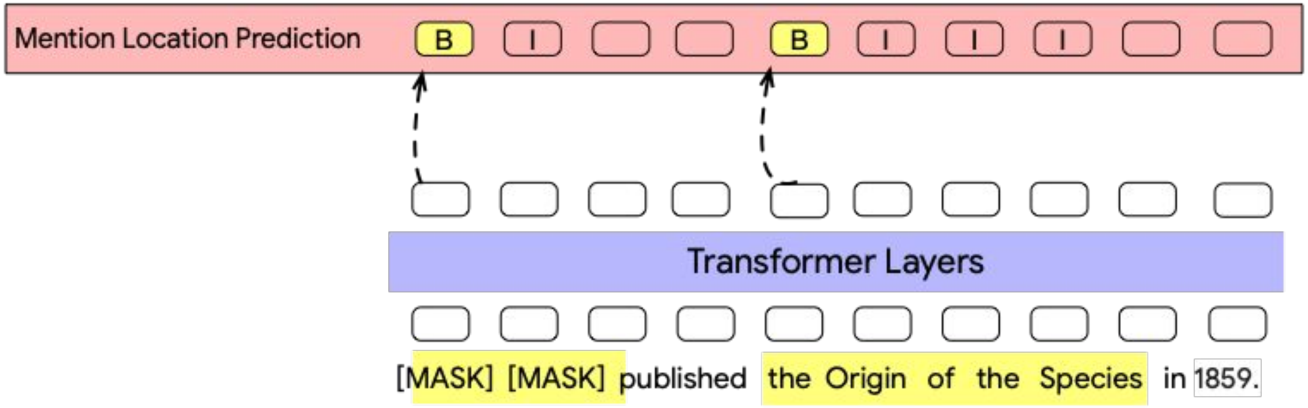
- Knowledge Enhanced Contextual Word Representations - Peters et.al. 2019
- ERNIE: Enhanced Language Representation with Informative Entities - Zhang et.al. 2019

EaE learns entity memories as part of the sequence model, rather than integrating pre-existing entity representations.



Transformer Layers

[MASK] [MASK] published the Origin of the Species in 1859.

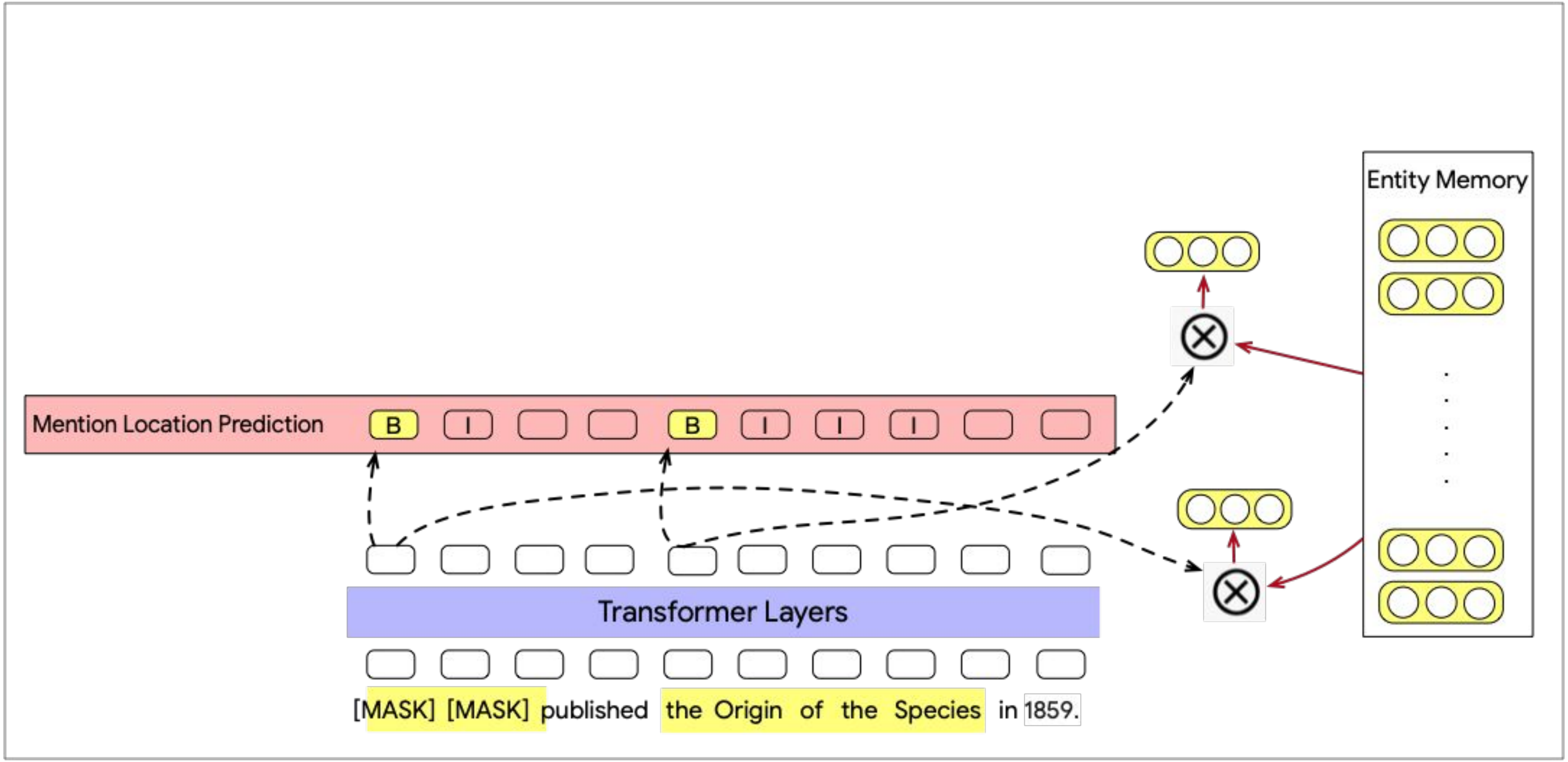


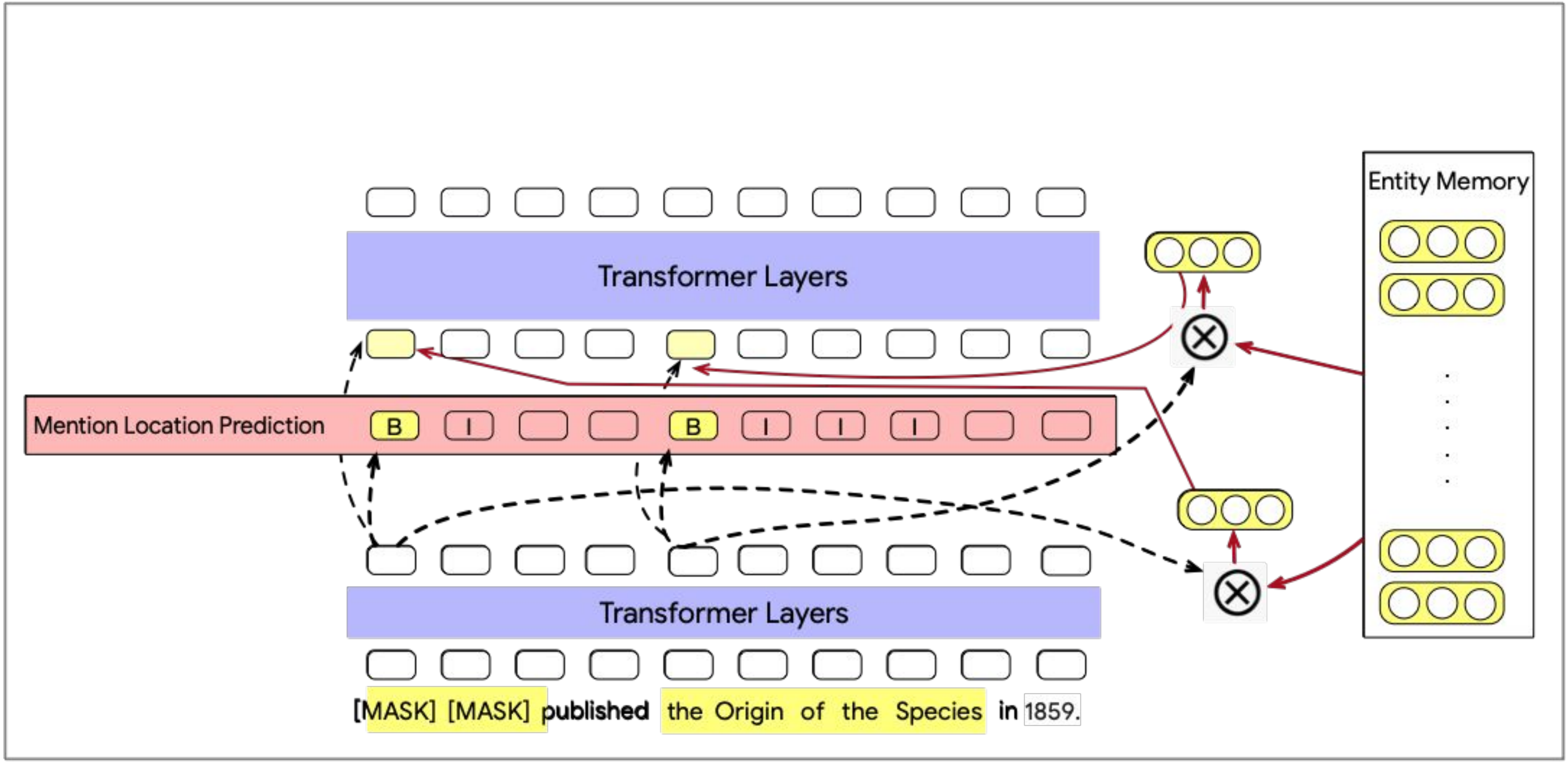
Mention Location Prediction

B I B I I I

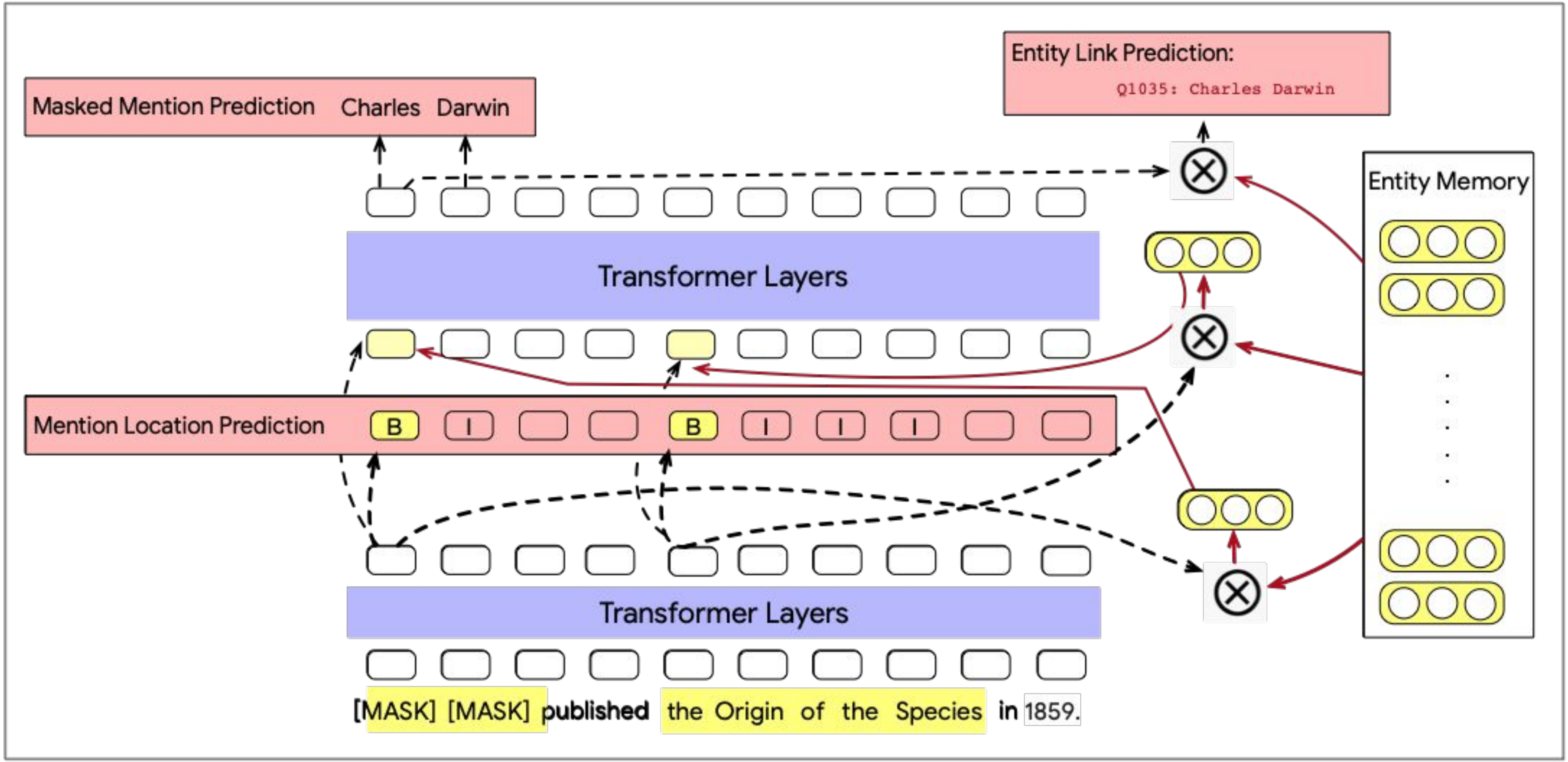
Transformer Layers

[MASK] [MASK] published the Origin of the Species in 1859.

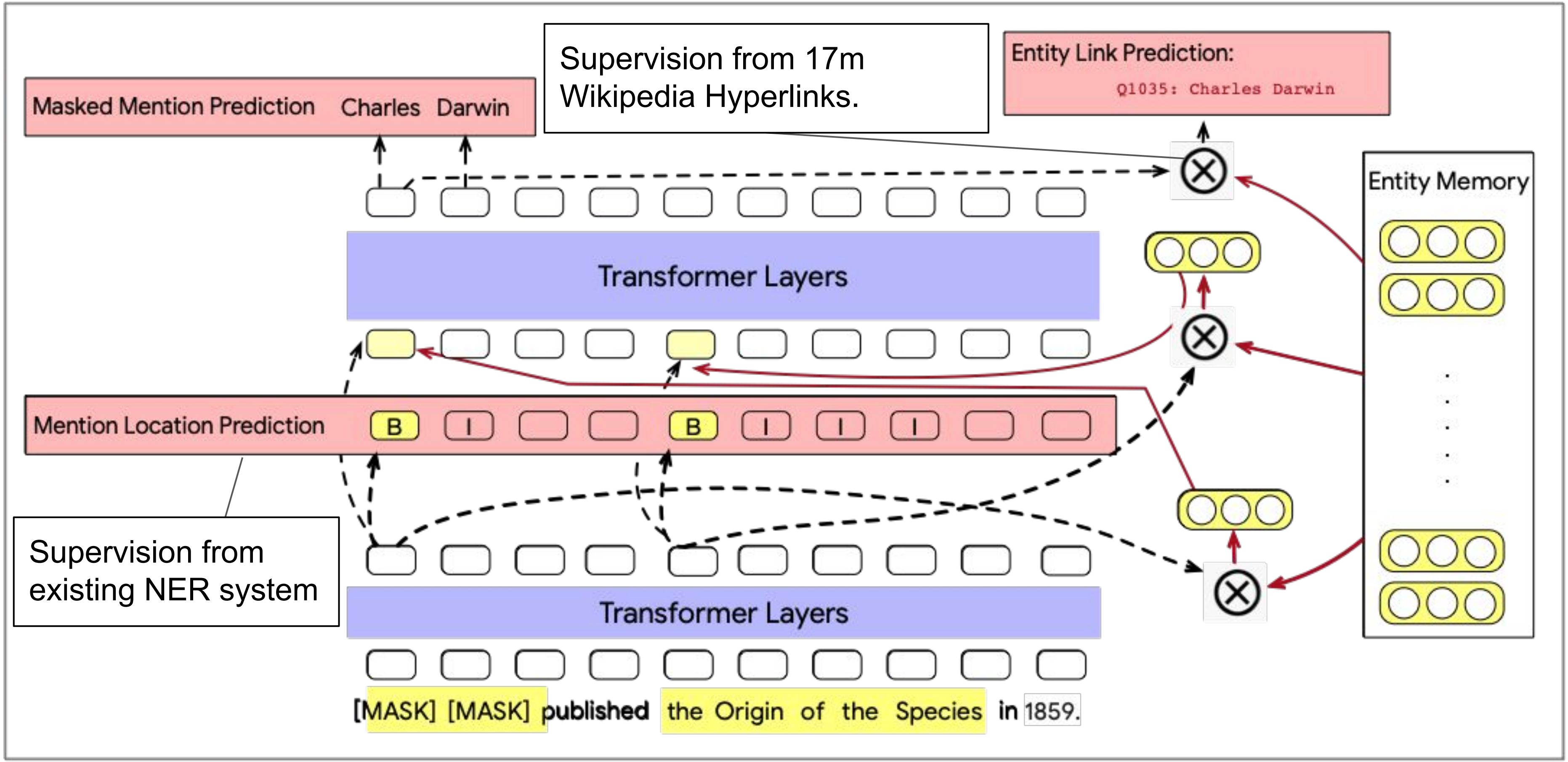

Entity Memory
Mention Location Prediction
B
I
B
I
I
I
Transformer Layers
[MASK] [MASK] published the Origin of the Species in 1859.



Google Research



Google Research



Supervision from 17m Wikipedia Hyperlinks.

Supervision from existing NER system

Evaluating EaE with knowledge probes: in the paper

Who directed the 2011 Palme d'Or winning film The Tree Of Life?

Open Domain Question Answering - TriviaQA

____ ____ published the Origin of the Species in 1859

Cloze prediction and link prediction - Wikipedia

The theory of relativity was developed by ____

Cloze prediction - LAMA - SQuAD

What team does Pudge Rodriguez play for?

Open Domain Question Answering - WebQuestions

Billy Mays, the undisputed king of TV yell and sell, died at his home in **Tampa**, Fla, on Sunday. **per:city_of_death**

Relation Extraction - TACRED

Adolphe Adam died in ____

Cloze prediction - LAMA - RE

person

They have been asked to appear in court to face the charge.

Entity Typing

Time is ____

Cloze prediction - LAMA - ConceptNet

Joe Cocker is represented by music label ____

Cloze prediction - LAMA - T-Rex

Evaluating EaE with knowledge probes: in this talk

Who directed the 2011 Palme d'Or winning film The Tree Of Life?

Open Domain Question Answering - TriviaQA

____ ____ published the Origin of the Species in 1859

Cloze prediction and link prediction - Wikipedia

The theory of relativity was developed by ____

Cloze prediction - LAMA - SQuAD

What team does Pudge Rodriguez play for?

Open Domain Question Answering - WebQuestions

Billy Mays, the undisputed king of TV yell and sell, died at his home in **Tampa**, Fla, on Sunday. **per:city_of_death**

Relation Extraction - TACRED

Adolphe Adam died in ____

Cloze prediction - LAMA - RE

person

They have been asked to appear in court to face the charge.

Entity Typing

Time is ____

Cloze prediction - LAMA - ConceptNet

Joe Cocker is represented by music label ____

Cloze prediction - LAMA - T-Rex

Google Research

What is the best way to store and retrieve facts about entities?

1. Is it better to mask wordpieces (Charl##) or mentions (Charles Darwin)?
2. Do we need every parameter for every example?
3. How important is providing entity memory supervision with Wikipedia hyperlinks?

Architecture	Masking strategy	Memory supervision	SQuAD	T-Rex
BERT Large *366M transformer params*	word-piece	NA	17.4	32.3
	mention	NA	24.4	31.4
EaE *110m transformer params* *256M memory params*	mention	NA	23.1	30.0
	mention	entity linking	22.4	37.4

Google Research

What is the best way to store and retrieve facts about entities?

1. Is it better to mask wordpieces (Charl##) or mentions (Charles Darwin)?
2. Do we need every parameter for every example?
3. How important is providing entity memory supervision with Wikipedia hyperlinks?

Architecture	Masking strategy	Memory supervision	SQuAD	T-Rex
BERT Large *366M transformer params*	word-piece	NA	17.4	32.3
	mention	NA	24.4	31.4
EaE *110m transformer params* *256M memory params*	mention	NA	23.1	30.0
	mention	entity linking	22.4	37.4

What is the best way to store and retrieve facts about entities?

1. Is it better to mask wordpieces (Charl##) or mentions (Charles Darwin)?
2. Do we need every parameter for every example?
3. How important is providing entity memory supervision with Wikipedia hyperlinks?

Architecture	Masking strategy	Memory supervision	SQuAD	T-Rex
BERT Large *366M transformer params*	word-piece	NA	17.4	32.3
	mention	NA	24.4	31.4
EaE *110m transformer params* *256M memory params*	mention	NA	23.1	30.0
	mention	entity linking	22.4	37.4

Google Research

What is the best way to store and retrieve facts about entities?

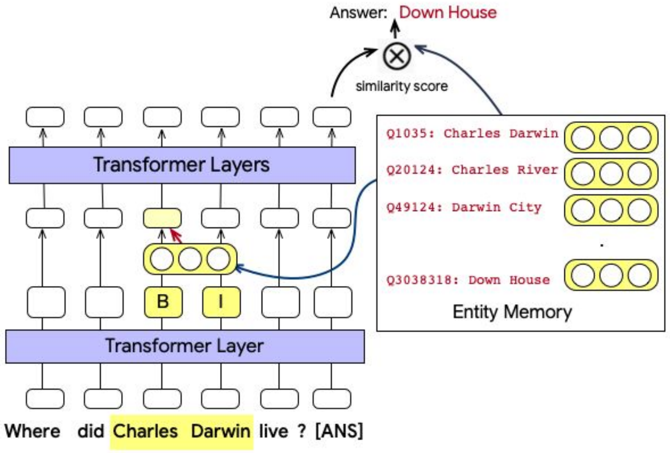
1. Is it better to mask wordpieces (Charl##) or mentions (Charles Darwin)?

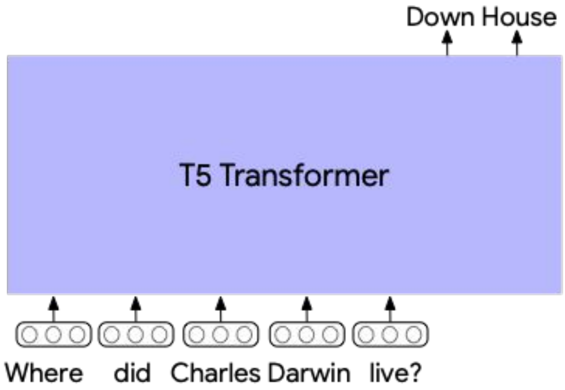
2. Do we need every parameter for every example?

3. How important is providing entity memory supervision with Wikipedia hyperlinks?

Architecture	Masking strategy	Memory supervision	SQuAD	T-Rex
BERT Large *366M transformer params*	word-piece	NA	17.4	32.3
	mention	NA	**24.4**	31.4
EaE *110m transformer params* *256M memory params*	mention	NA	23.1	30.0
	mention	entity linking	22.4	**37.4**

Open Domain Question Answering Types

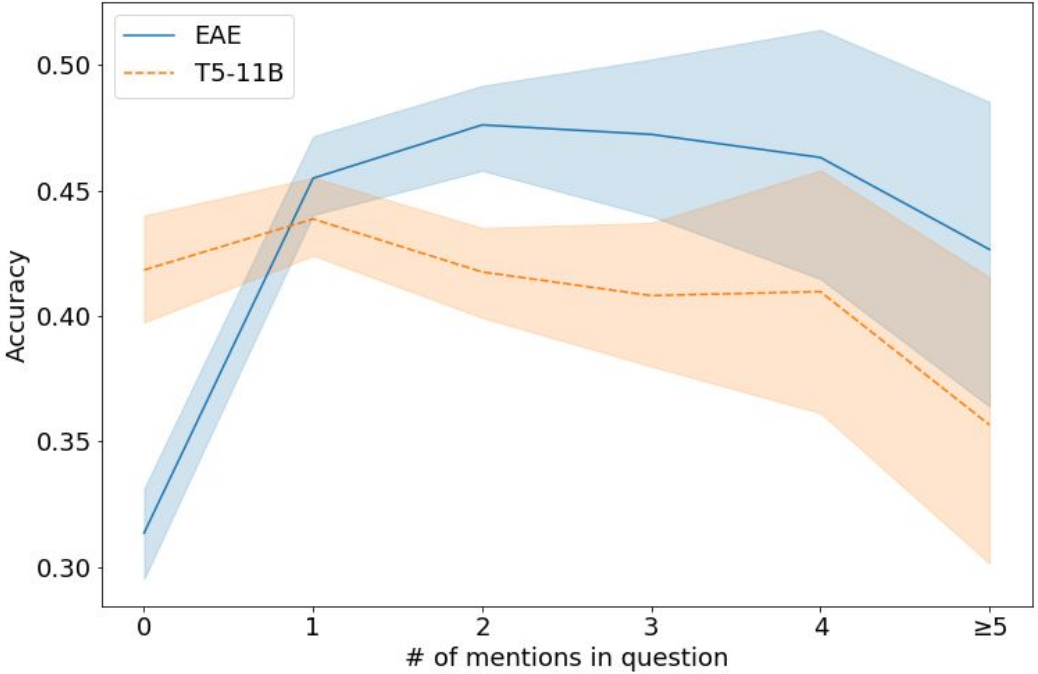




TriviaQA Closed Book Question Answering

	Parameters		**Results**	
	All	**Activated**	**TriviaQA**	**WebQ**
T5 3B	3B	3B	35.1	33.6
T5 11B	11B	11B	42.3	37.4
T5 11B + Salient Span Masking	11B	11B	53.3	43.5
EaE	367M	95M	43.2	39.0

Analysis: sensitivity to entities in the question



Qualitative Analysis: Predictions on TriviaQA

Q: Next Sunday, Sept 19, is International what day?

A: Talk like a pirate day

T5: ✓

EaE: Pearl Harbor Remembrance Day

Qualitative Analysis: Predictions on TriviaQA

Q: Which **Dr. Who** villain has been played by **Roger Delgado**, **Anthony Ainley**, **Eric Roberts**, etc?

A: The Master

T5: mr. daleks

EaE: ✓

Google Research

Qualitative Analysis: Predictions on TriviaQA

Q: Which early aviator flew in a plane christened **Jason**?

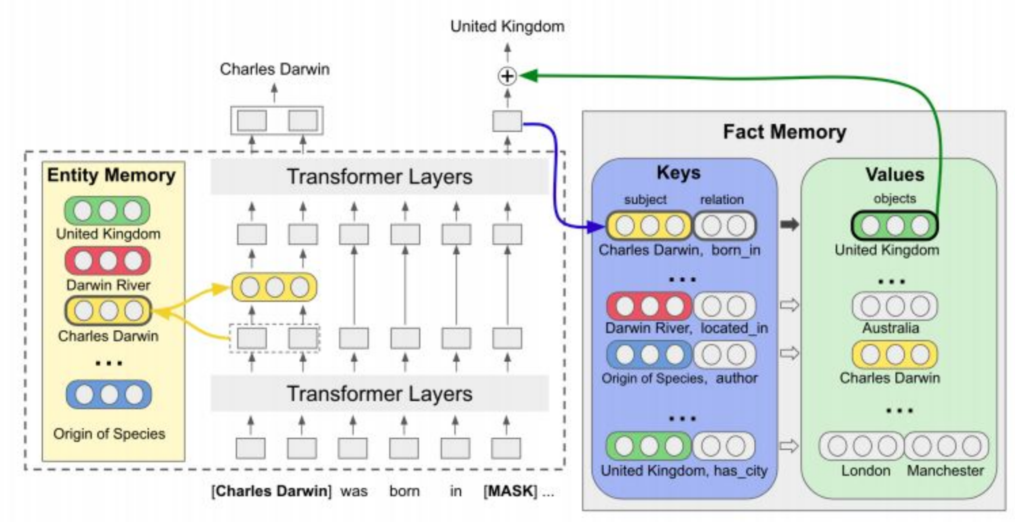
A: Amy Johnston

T5: jean batten

EaE: Icarus, Jason linked to Jason (Greek Mythology)

More to come in this direction

Facts as Experts adds a fact memory to EaE



Google Research

Teaser: Entity Discovery

Working on combining representation learning with new, scalable clustering methods for work on Entity Discovery

Initial joint work with Rob Logan (UC Irvine), Sameer Singh (UC Irvine), Andrew McCallum (Google).

Continuing work with Andrew McCallum.

Wrapping up

Entities: grounded form of information

Useful abstractions, but where are they?

- we can bootstrap grounding by analyzing text
- we can exploit inherent multilingual landscape to bring together a much more comprehensive set of entities

How can we model and use them?

- dual encoders without alias tables
- cross-attention models
- incorporation into Transformer-based downstream models

THANK YOU!