RICE[1]

Microsoft[2]

Instructed Language Models with Retrievers Are Powerful Entity Linkers

Zilin Xiao[1], Ming Gong[2], Jie Wu[2],
Xingyao Zhang[2], Linjun Shou[2], Daxin Jiang[2]

Rice University[1], Microsoft STCA[2]

EMNLP 2023

Paper Video

Task: Entity Linking (EL)

Entity Linking is the process of grounding entity mentions in the document to an external knowledge base (e.g. Wikipedia entity).

After the death of **Steve** , the former CEO of **Apple** ...

Steve_Jobs *Apple_Inc.*

Steve dropped out of **Stanford** to join **Microsoft...**

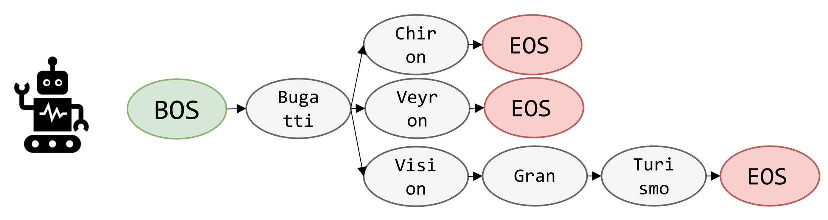
Steve_Ballmer *Stanford_University* *Microsoft*

Previous Generative EL: EL as Constrained Decoding

Document

After the death of Steve, the former CEO of Apple ...

Sequence-to-sequence Model



Trie-constrained Decoding ensures a legal suquence that can be parsed.

Generated

After the death of Steve, the former CEO of [Apple] {Apple Inc.} ...

Major Drawback: Have to call LM at each document token!

InsGenEL: Retrieval-augmented Generative EL

Document

After the death of Steve, the former CEO of Apple ...

Doc Encoder

Top-*K* Retrieval

Retrieved Entities

1. *Apple* (Q89)
2. *Apple Inc.* (Q312)
3. *Steve Jobs* (Q19837)
4. *United States* (Q30)
5. *Dell* (Q30873)
6. *Google* (Q95)

......

Possible Mentions

Apple (from *Apple*)
iOS (from *Apple*)
Apple (from *Apple Inc.*)
Steve (from *Steve Jobs*)
Jobs (from *Steve Jobs*)
CEO (from *Steve Jobs*)
Apple CEO (from *Steve Jobs*)
Dell (from *Dell*)
Google (from *Google*)
...

Surface Form Matching

Dynamic Mention-Candidate Mapping

Steve: ['**Steve Jobs**']; **CEO**: ['**Steve Jobs**'];
Apple: ['**Apple**', '**Apple Inc.**']

After the death of Steve, the former CEO of Apple...

______ : **Decision-required Spans**

Guided Entity Linking Generation

After the death of [Steve]{ *Steve Jobs* }, the former CEO of [Apple]{ *Apple* *Inc.* }

Steve (greedy decoding, thus do not continue)

call generative model
only when decision is needed

Generated

After the death of [Steve] {Steve Jobs}, the former CEO of [Apple] {Apple Inc.} ...

Number of Forward Passes: **4**

Generated

After the death of Steve, the former CEO of [Apple] {Apple Inc.} ...

Number of Forward Passes: **64** with beam size of 6

Training

After the death of Steve...

$$\mathcal{L}_{\mathrm{EL}} = -\sum_{i=n}^{N} \log P(y_i \mid y_1, \ldots, y_{i-1})$$

After the death of [Steve] {Steve Jobs}...

Language Modeling

After bowling [Somerset]$_3$ out for 83 on the opening morning at [Grace Road]$_2$, [Leicestershire]$_1$ extended their first innings by 94 runs before being bowled out for 296 with [England]$_{11}$

Top-*K* candidate entities

1. Leicestershire County Cricket Club
2. Grace Road
3. Somerset County Cricket Club
4. ✗ Durham County Cricket Club
5. ✗ Nottinghamshire County Cricket Club
6. ✗ Derbyshire County Cricket Club
7. ✗ Warwickshire County Cricket Club
8. ✗ Leicestershire
9. ✗ Worcestershire County Cricket Club
10. ✗ Yorkshire County Cricket Club
11. England cricket team
12. ✗ Marylebone Cricket Club
13. ✗ Sussex County Cricket Club
14. ✗ Kent County Cricket Club
15. ✗ Leicester
16. ✗ Aylestone Road
17. ✗ County Cricket Ground, Derby

⋮

During training, we prepare a document chunk x and a set of oracle entities $\mathcal{E}(x) \in \mathcal{E}$ that are mentioned in x (colorful entities). We train the retriever with maximizing the following objective:

$$\sum_{e \in \mathcal{E}(x)} \log \left(\frac{\exp(S(e))}{\exp(S(e)) + \sum_{e' \in \mathbf{N}(\mathcal{E},x)} \exp(S(e'))} \right)$$

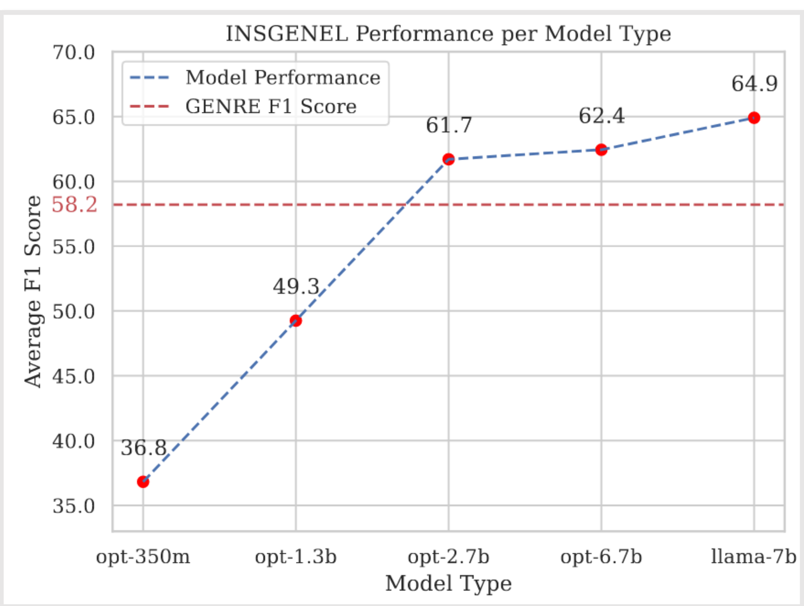
where $S(e) = X_1^\top E_1^e$ stands for the matching score between document chunk x and entity e, $\mathbf{N}(\mathcal{E}, x)$ is a set of negative entities (black entities) that do not overlap with gold entity set $\mathcal{E}(x)$.

Retriever Training

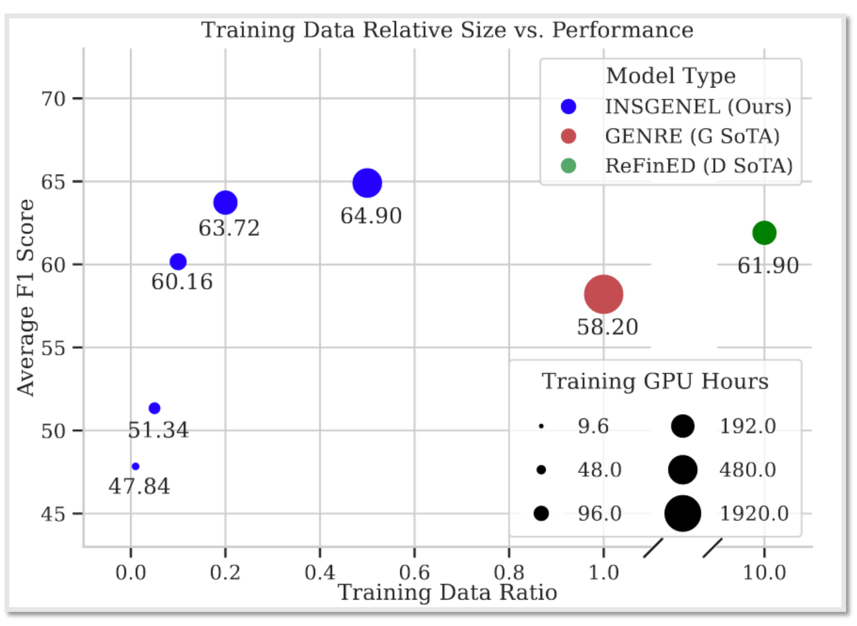
Results & Ablations

Category	Method	In-domain AIDA	MSNBC	Der	K50	R128	R500	OKE15	OKE16	Avg
Discriminative	Hoffart et al. (2011)	72.8*	65.1	32.6	55.4	46.4	42.4	63.1	0.0	47.2
	Kolitsas et al. (2018)	82.4*	72.4	34.1	35.2	50.3	38.2	61.9	52.7	53.4
	van Hulst et al. (2020)	80.5*	72.4	41.1	50.7	49.9	35.0	63.1	58.3	56.4
	Zhang et al. (2022b)	**85.8***	72.1	52.9	64.5	54.1	41.9	61.1	51.3	60.5
	Ayoola et al. (2022)	84.0*	71.8	50.7	64.7	58.1	42.0	64.4	59.1	61.9
Generative	Cao et al. (2021b)	83.7*	73.7	54.1	60.7	46.7	40.3	56.1	50.0	58.2
	Cao et al. (2021a)	85.5*	-	-	-	-	-	-	-	-
	Mrini et al. (2022b)	85.7*	-	-	-	-	-	-	-	-
Ours	InsGenEL	81.5	69.5	**60.9**	**73.8**	**58.6**	**46.8**	**65.7**	62.1	**64.9**
	InsGenEL-R	80.6	**74.2**	59.8	71.9	56.8	45.5	64.1	**63.3**	64.5

Out-of-domain: MSNBC, Der, K50, R128, R500, OKE15, OKE16.



Model Type and Size



Training Efficiency (Compute & Data)

Exploration: Can we prompt GPT-3 do generative EL in zero-shot setting?

```
### Your task is to read the example document and
identify the mentions…

### Rules
…
## Test Document
After the death of Steve, the former CEO of Apple, his
commencement speech at Stanford was watched thousands of
times.
## Test Document Candidates
- candidate 0 for test document: Stanford University
- candidate 1 for test document: Apple Inc.
- …

## Answers
```

Method	AIDA	MSNBC	K50	R500	Avg
InsGenEL-ICL	-	-	-	-	-
- text-davinci-003	50.0	53.3	39.2	34.9	44.4
- code-davinci-002	60.7	47.4	39.0	25.4	43.1
InsGenEL-R	80.6	74.2	71.9	45.5	68.1

Far from satisfactory!

Takeaways

1. InsGenEL makes Generative EL more efficient and accurate.
2. EL still remains a persistent hurdle for GPT-3 even with heavy prompt engineering and retrieved candidates.

Reference

Wenzheng Zhang, Wenyue Hua, and Karl Stratos. Entqa: Entity linking as question answering. ICLR 2022.
Nicola De Cao, Gautier Izacard, Sebastian Riedel, and Fabio Petroni. Autoregressive entity retrieval. ICLR 2021