



Unified Examination of Entity Linking in Absence of Candidate Sets

NAACL 2024



Nicolas Ong



Hassan S. Shavarani



Anoop Sarkar

Natural Language Processing Lab (Natlang)
School of Computing Science
Simon Fraser University

June 12, 2024

Outline

Introduction

Unified Evaluation of Entity Linking Systems

Candidate Set Ablations

Results & Analysis

Summary

Entity Linking

Definition

To annotate plain text with references to entities from a knowledge base.

“Paris is the capital of France”

wikipedia.org/wiki/Paris

wikipedia.org/wiki/France

Entity Linking

Candidate Sets

Common technique: Candidate Sets.

Reduces candidate entities to a manageable amount (Wikipedia has $\sim 5M$)

Candidate Sets

Example

- ▶ **Apple** is an American multinational corporation and technology company headquartered in Cupertino, California, in Silicon Valley.
- ▶ An **apple** is a round, edible fruit produced by an apple tree.

```
['Apple_Inc.', 'Apple', 'Apple_Records',  
'Application_store', 'Apple_Corps', 'Apple_(band)',  
'Apple_Store', 'Apple_(album)', 'Apple,_Oklahoma',  
  'Fiona_Apple', 'Apple_(automobile)',  
'Apple_(symbolism)', 'Michael_Apple', 'Thomas_Apple',  
  'Martin_Apple', 'Max_Apple', 'Terminal_(OS_X)',  
  'Billy_Apple', 'Apple_Dylan', 'Chris_Apple',  
  'Jim_Apple', 'Henry_Harbaugh_Apple',  
'Apple_Network_Server', 'Apple_Hong', 'Pat_Apple',  
  'Apple_Comics', 'Apple_Pope']
```

Motivation

Many ontologies and languages lack the high quality candidate sets used by many modern entity linking systems.

Are these systems resilient to the absence of candidate sets? →
Candidate set ablation experiments

Unified Evaluation

Environment

To compare different entity linking systems, we need to evaluate them in a unified black-box evaluation environment.

Our environment consists of of GERBIL and `gerbil_connect`.

Unified Evaluation

Results

	Micro-F1			Difference		Modifications
	testa	testb	testc	testa	testb	
Kolitsas et al. (2018)	89.50	82.44	65.75	+0.10	+0.04	-
Peters et al. (2019)						1, 2, 3
KnowBert-Wiki	76.74	71.68	54.12	-3.46	-2.72	
KnowBert-W+W	77.19	71.69	53.92	-4.91	-2.01	
Poerner et al. (2020)	89.40	84.83	65.93	-1.40	-0.17	1, 2, 3, 5
van Hulst et al. (2020)						-
Wiki 2014	83.30	82.53	71.69	-	-0.77	
Wiki 2019	79.64	80.10	73.54	-	-0.40	
De Cao et al. (2021b)	90.09	82.78	75.60	-	-0.92	2, 4
De Cao et al. (2021a)	87.29	85.65	47.54	-	+0.15	4
Zhang et al. (2022)	86.81	84.30	72.55	-	-1.50	-
Feng et al. (2022)	87.64	86.49	65.05	-	+0.19	1, 2, 3, 5
Shavarani and Sarkar (2023)						-
large-500K (no cnds.)	89.72	82.25	77.54	+0.02	+0.05	
large-500K (Kb+Yago)	89.89	82.88	59.50	+0.09	+0.08	
large-500K (PPRforNED)	91.58	85.22	46.98	+0.08	+0.02	

Unified Evaluation

Modifications

1. Tokenization for input documents
2. Document splitting for input documents
3. Token-to-character resolution for output annotations
4. Creation of custom data
5. Training of a new model

Future authors should make their source code, trained models, and a `gerbil_connect` integration with GERBIL publicly available. At minimum, provide a function that accepts raw text and outputs a list of character-level annotations.

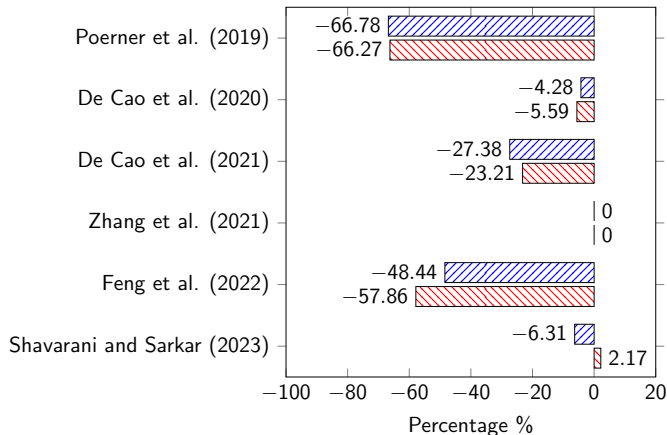
Candidate Set Ablations

Results

	Micro-F1		
	testa	testb	testc
De Cao et al. (2021b)	85.15	78.98	75.62
De Cao et al. (2021a)	62.00	49.51	37.05
Zhang et al. (2022)	86.81	84.30	72.55
Shavarani and Sarkar (2023)	89.72	82.25	77.54
Poerner et al. (2020)	22.81	18.81	17.56
Feng et al. (2022)	35.00	32.58	27.48

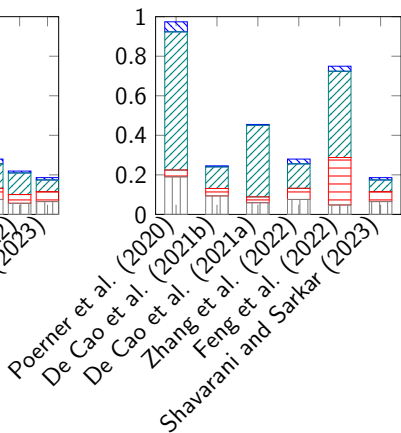
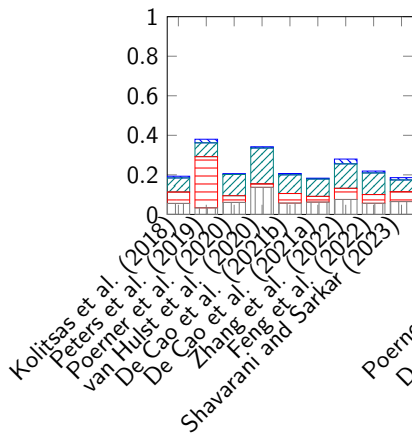
Results & Analysis

Impact on Precision and Recall



Results & Analysis

Impact on Errors Made



Results & Analysis

Conclusions

- ▶ Current entity linkers are very dependent on candidate sets.
- ▶ Generative and structured prediction approaches are robust.
- ▶ Large candidate sets increase run times for many systems.
- ▶ Candidate sets primarily decrease incorrect entity errors.

Summary

Contributions

- ▶ We benchmark end-to-end entity linking systems from the literature on the CoNLL/AIDA dataset in a unified evaluation environment consisting of GERBIL and `gerbil_connect`.
- ▶ We evaluate the resilience of these systems in the absence of pre-computed candidate sets using our unified evaluation environment. This environment also enables future ablation studies.
- ▶ We examine the adaptability of these systems to unseen test data using the novel AIDA/testc dataset.