technische universität
dortmund

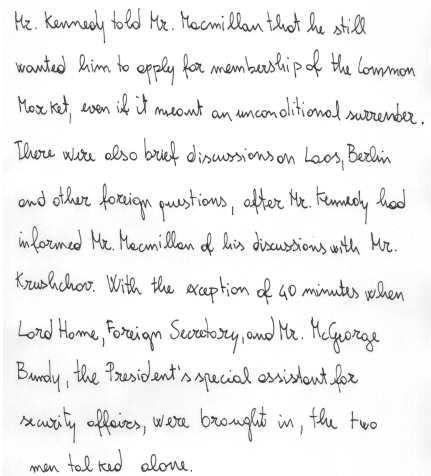
Named Entity Linking on Handwritten Document Images

Oliver Tüselmann and Gernot A. Fink
Department of Computer Science
TU Dortmund University
Dortmund, Germany

May 23th, 2022

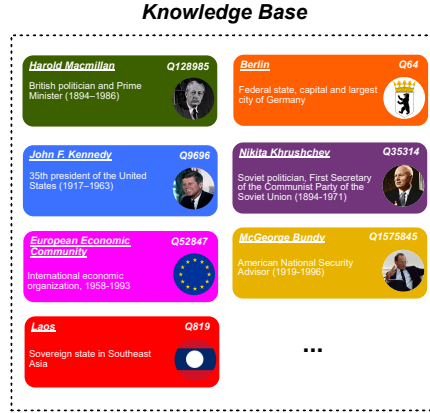
- Motivation
- Named Entity Linking
- Datasets
- Baseline Approach
- Results
- Conclusions

Motivation


Mr. Kennedy told Mr. Macmillan that he still wanted him to apply for membership of the Common Market, even if it meant an unconditional surrender. There were also brief discussions on Laos, Berlin and other foreign questions, after Mr. Kennedy had informed Mr. Macmillan of his discussions with Mr. Krushchov. With the exception of 40 minutes when Lord Home, Foreign Secretary, and Mr. McGeorge Bundy, the President's special assistant for security affairs, were brought in, the two men talked alone.

Motivation

Mr. Kennedy Q9696 told Mr. Macmillan Q128985 that he still wanted him to apply for membership of the Common Market Q52847 Q52847 even if it meant an unconditional surrender. There were also brief discussions on Laos Q819 Berlin Q64 and other foreign questions, after Mr. Kennedy Q9696 had informed Mr. Macmillan Q128985 of his discussions with Mr. Khrushchov Q35314. With the exception of 40 minutes when Lord Home, Foreign Secretary, and Mr. McGeorge Q1575845 Bundy Q1575845, the President's special assistant for security affairs, were brought in, the two men talked alone.


Knowledge Base
Harold Macmillan Q128985
British politician and Prime Minister (1894–1986)
Berlin Q64
Federal state, capital and largest city of Germany
John F. Kennedy Q9696
35th president of the United States (1917–1963)
Nikita Khrushchev Q35314
Soviet politician, First Secretary of the Communist Party of the Soviet Union (1894-1971)
European Economic Community Q52847
International economic organization, 1958-1993
McGeorge Bundy Q1575845
American National Security Advisor (1919-1996)
Laos Q819
Sovereign state in Southeast Asia
...

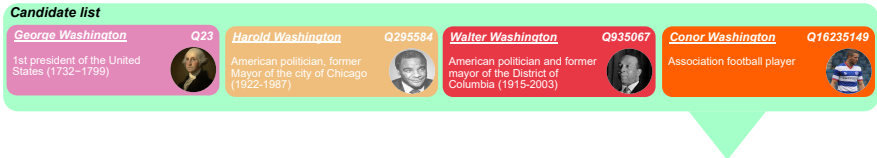
Example

Many places and monuments have been named in honour of Washington, most notably the capital of the United States, Washington, D.C. In addition, Washington is the only US state named after a president.

1.Step: Mention Detection

Many places and monuments have been named in honour of Washington PERSON ,
most notably the capital of the United States LOCATION , Washington, D.C. LOCATION
In addition, Washington LOCATION is the only US LOCATION state named after a president.

2.Step: Candidate Selection

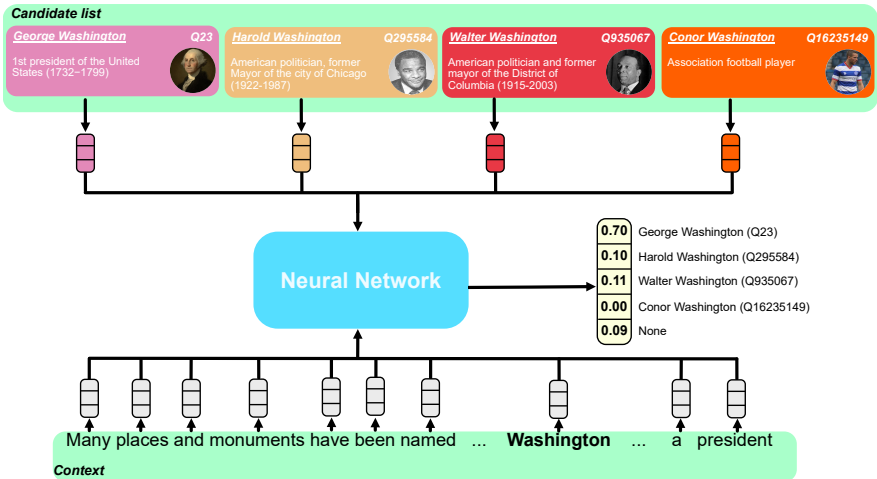


Many places and monuments have been named in honour of Washington PERSON ,

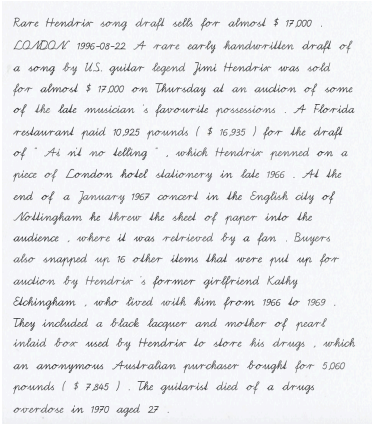
most notably the capital of the United States LOCATION , Washington, D.C. LOCATION

In addition, Washington LOCATION is the only US LOCATION state named after a president.

3.Step: Entity Ranking

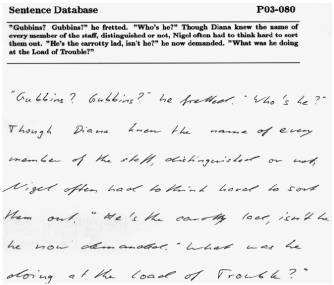


Datasets



Rare Hendrix song draft sells for almost $ 17,000 .
LONDON 1996-08-22 A rare early handwritten draft of a song by U.S. guitar legend Jimi Hendrix was sold for almost $ 17,000 on Thursday at an auction of some of the late musician 's favourite possessions . A Florida restaurant paid 10,925 pounds ($ 16,935) for the draft of " At ni no telling " , which Hendrix penned on a piece of London hotel stationery in late 1966 . At the end of a January 1967 concert in the English city of Nottingham he threw the sheet of paper into the audience , where it was retrieved by a fan . Buyers also snapped up 16 other items that were put up for auction by Hendrix 's former girlfriend Kathy Etchingham , who lived with him from 1966 to 1969 . They included a black lacquer and mother of pearl inlaid box used by Hendrix to store his drugs , which an anonymous Australian purchaser bought for 5,060 pounds ($ 7,845) . The guitarist died of a drugs overdose in 1970 aged 27 .

(a) HW-AIDA-CoNLL

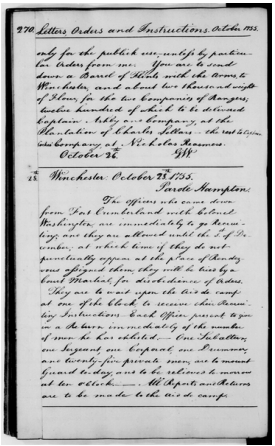


Sentence Database **P03-080**

"Gubbins? Gubbins?" he fretted. "Who's he?" Though Diana knew the name of every member of the staff, distinguished or not, Nigel often had to think hard to sort them out. "He's the carroty lad, isn't he?" he now demanded. "What was he doing at the Load of Trouble?"

"Gubbins? Gubbins?" he fretted. "Who's he?" Though Diana knew the name of every member of the staff, distinguished or not, Nigel often had to think hard to sort them out. "He's the carroty lad, isn't he?" he now demanded. "What was he doing at the Load of Trouble?"

(b) IAM-DB



270 Letters Orders and Instructions. October 1755.

only for the publick use, unless by particular Orders from me. You are to send down a Board of Field with the Arms to Winchester, and about two thousand weight of Flour for the two Companys of Rangers, twelve hundred of which to be delivered Captain Ashby's Company at the Plantation of Charles Sellers the rest to Captain Lovelapp's at Nicholas Rosemans.

October 24. 1755

28 Winchester October 25th 1755.

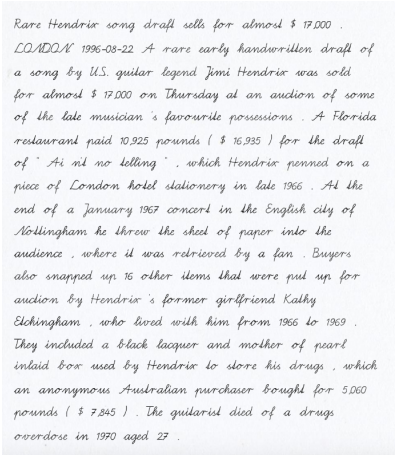
Parole Hampton.

The Officers who came down from Fort Cumberland with Colonel Washington are immediately to go down to their Companys and are allowed until the 1st of November, at which time if they do not punctually appear at the place of Rendezvous appointed them, they will be tried by a Court Martial for disobedience of Orders.

They are to wait upon the Field Officer of the Day at Guard to receive Instructions. Each Officer present upon a Review, or inspection of the Arms of his Company shall be fined for the first offence Twenty Shillings, for the second Forty Shillings, and for the third he shall be cashiered. Captain Lovelapp's Company, Captain Ashby's Company, and twenty-five private men are to mount Guard this day, and to be relieved to morrow at ten o'Clock. All Reports and Returns are to be made to the Field Camp.

(c) George Washington

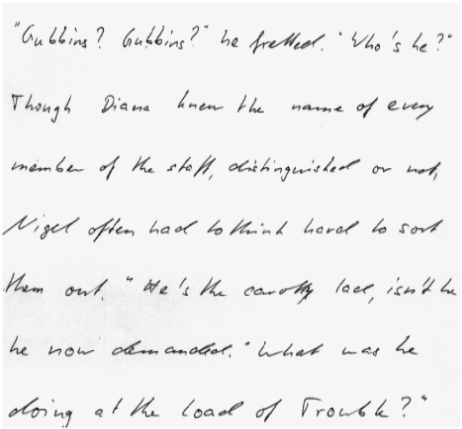
HW-AIDA-CoNLL

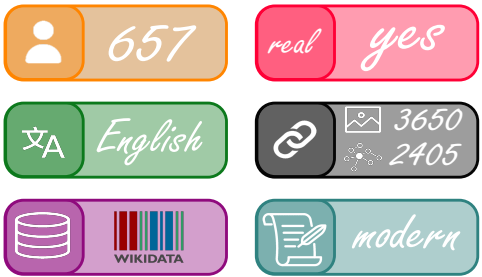


Rare Hendrix song draft sells for almost $ 17,000 .
LONDON 1996-08-22 A rare early handwritten draft of a song by U.S. guitar legend Jimi Hendrix was sold for almost $ 17,000 on Thursday at an auction of some of the late musician 's favourite possessions . A Florida restaurant paid 10,925 pounds ($ 16,935) for the draft of " Ai nt no telling " , which Hendrix penned on a piece of London hotel stationery in late 1966 . At the end of a January 1967 concert in the English city of Nottingham he threw the sheet of paper into the audience , where it was retrieved by a fan . Buyers also snapped up 16 other items that were put up for auction by Hendrix 's former girlfriend Kathy Etchingham , who lived with him from 1966 to 1969 . They included a black lacquer and mother of pearl inlaid box used by Hendrix to store his drugs , which an anonymous Australian purchaser bought for 5,060 pounds ($ 7,845) . The guitarist died of a drugs overdose in 1970 aged 27 .

[1] J. Hoffart, M. A. Yosef, I. Bordino, H. Fürstenau, M. Pinkal, M. Spaniol, B. Taneva, S. Thater . G. Weikum: *Robust Disambiguation of Named Entities in Text*. . *Proc. Conf. on Empirical Methods in Natural Language Processing*, . 782–792, Edinburgh, UK, 2011

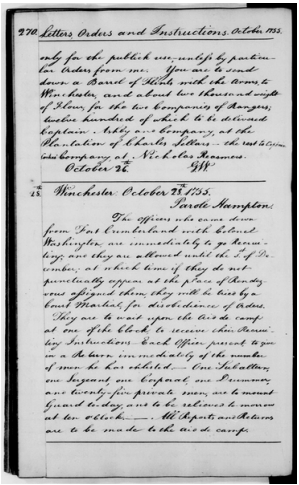
IAM-DB





[1] U. Marti . H. Bunke: *The IAM-database: An English sentence database for offline handwriting recognition.* Int. Journal on Document Analysis and Recognition, 5(1):39–46, 2002

George Washington


270. Letters Orders and Instructions. October 1755.

only for the publick use—unless by particular Orders from me. You are to send down a Barrel of Flints with the Arms to Winchester, and about two thousand weight of Flour, for the two Companies of Rangers; twelve hundred of which to be delivered Captain Ashby and Company, at the Plantation of Charles Sellars—the rest to Captain Cocke's Company at Nicholas Reamers.

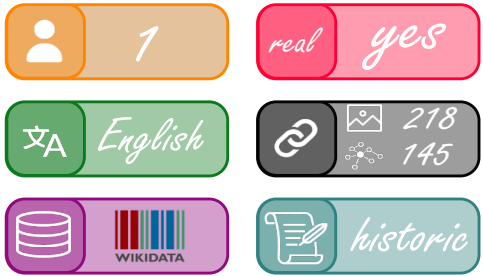
October 26th. G.W.

28th Winchester October 28th 1755.

Parole Hampton.

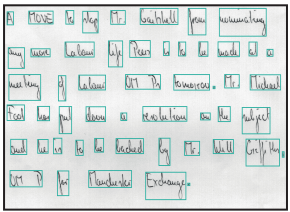
The officers who came down from Fort Cumberland with Colonel Washington, are immediately to go Recruiting; and they are allowed until the 5th of December; at which time if they do not punctually appear at the place of Rendezvous assigned them, they will be tried by a Court Martial, for disobedience of Orders.

They are to wait upon the Aid de camp at one of the Clock, to receive their Recruiting Instructions—Each Officer present to give in a Return immediately of the number of men he has obtained.—One Subaltern, one Sergeant, one Corporal, one Drummer, and twenty-five private men, are to mount Guard to-day, and to be relieved to morrow at ten o'Clock.—All Reports and Returns are to be made to the Aid de camp.



[1] T. M. Rath . R. Manmatha: *Word spotting for historical documents*. Int. Journal on Document Analysis and Recognition, 9(2-4):139–152, 2007

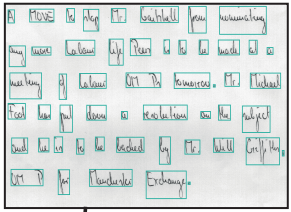
Baseline Approach



Knowledge Base

Hugh Gaitskell Q33738	Labour Party Q9630
William Griffith Q8010224	Michael Foot Q332419
Manchester Exchange Q11797268	

Baseline Approach



Knowledge Base

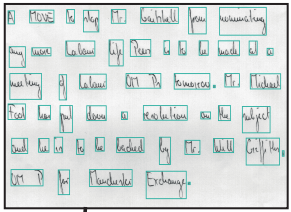
Hugh Gaitskell Q33738	Labour Party Q9630
William Griffith Q8010224	Michael Foot Q332419
Manchester Exchange Q11797268	

Word Image Handwriting Recognizer [16]

A MOVE to stap Mr. Gaihkell from nominating ang move Labour lief Perrs in to he made at a meeting of Labour OM Pr tomorrou . Mr. Mickael Foot her put down a resolution on the subject and he is to be backed by Mr. Will Griffiths , OM P fro Manchester Exchange .

[16] L. Kang, J. I. Toledo, P. Riba, M. Villegas, A. Fornés . M. Rusiñol: *Convolve, Attend and Spell: An Attention-based Sequence-to-Sequence Model for Handwritten Word Recognition*. . *Proc. German Conf. on Pattern Recognition*, . 459–472, Stuttgart, Germany, 2018

Baseline Approach



Knowledge Base

Hugh Gaitskell Q33738	Labour Party Q9630
William Griffith Q8010224	Michael Foot Q332419
Manchester Exchange Q11797268	

two-staged approach

Word Image Handwriting Recognizer [16]

Textual NEL Model [15]

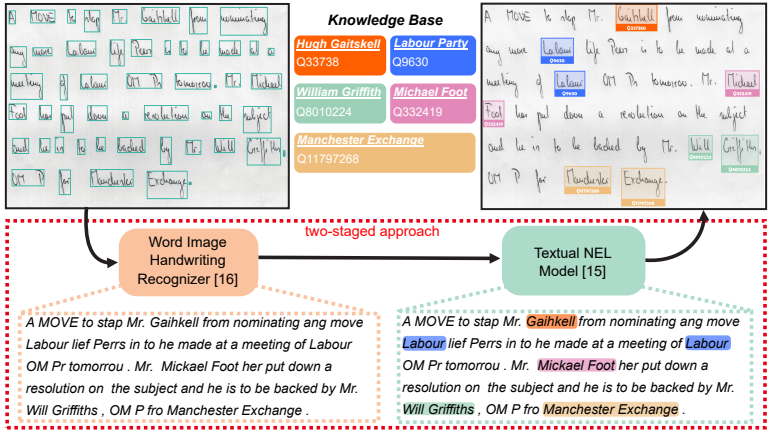
A MOVE to stap Mr. Gaihkell from nominating ang move Labour lief Perrs in to he made at a meeting of Labour OM Pr tomorrou . Mr. Mickael Foot her put down a resolution on the subject and he is to be backed by Mr. Will Griffiths , OM P fro Manchester Exchange .

A MOVE to stap Mr. Gaihkell from nominating ang move Labour lief Perrs in to he made at a meeting of Labour OM Pr tomorrou . Mr. Mickael Foot her put down a resolution on the subject and he is to be backed by Mr. Will Griffiths , OM P fro Manchester Exchange .

[16] L. Kang, J. I. Toledo, P. Riba, M. Villegas, A. Fornés . M. Rusiñol: *Convolve, Attend and Spell: An Attention-based Sequence-to-Sequence Model for Handwritten Word Recognition*. . *Proc. German Conf. on Pattern Recognition*, . 459–472, Stuttgart, Germany, 2018

[15] J. M. van Hulst, F. Hasibi, K. Dercksen, K. Balog . A. P. de Vries: *REL: An Entity Linker Standing on the Shoulders of Giants*. . *Proc. Int. ACM SIGIR Conf. on Research and Development in Information Retrieval*, . 2197–2200, Xi'an, China, 2020

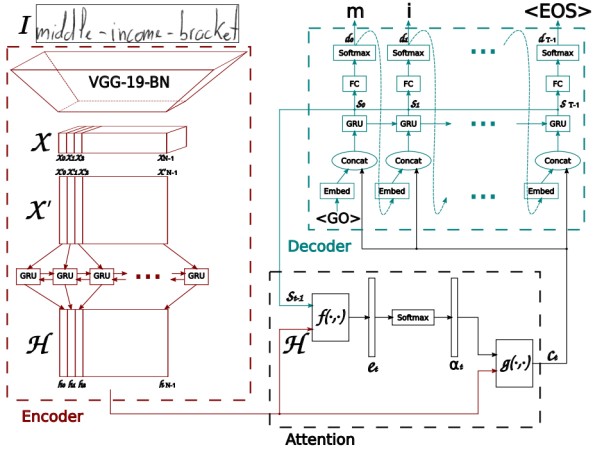
Baseline Approach



[16] L. Kang, J. I. Toledo, P. Riba, M. Villegas, A. Fornés . M. Rusiñol: *Convolve, Attend and Spell: An Attention-based Sequence-to-Sequence Model for Handwritten Word Recognition*. . *Proc. German Conf. on Pattern Recognition*, . 459–472, Stuttgart, Germany, 2018

[15] J. M. van Hulst, F. Hasibi, K. Dercksen, K. Balog . A. P. de Vries: *REL: An Entity Linker Standing on the Shoulders of Giants*. . *Proc. Int. ACM SIGIR Conf. on Research and Development in Information Retrieval*, . 2197–2200, Xi'an, China, 2020

Handwriting Text Recognizer



[1] L. Kang, J. I. Toledo, P. Riba, M. Villegas, A. Fornés . M. Rusiñol: *Convolve, Attend and Spell: An Attention-based Sequence-to-Sequence Model for Handwritten Word Recognition*. . *Proc. German Conf. on Pattern Recognition*, . 459–472, Stuttgart, Germany, 2018

REL: Radbound Entity Linker

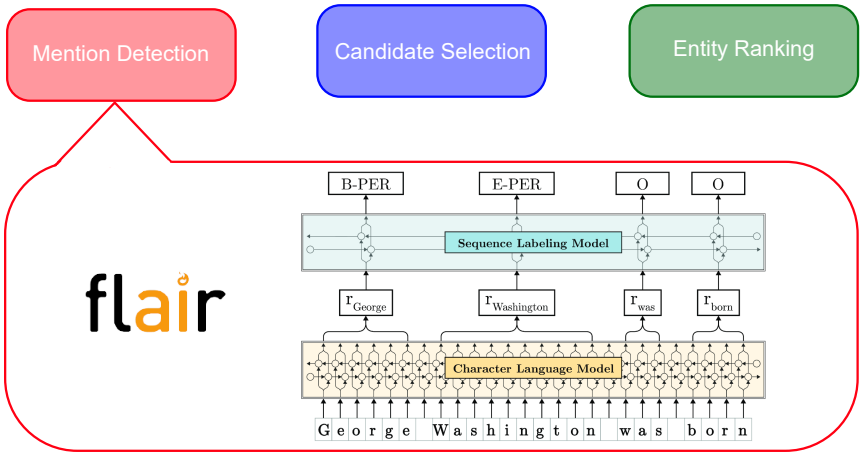
[1] J. M. van Hulst, F. Hasibi, K. Dercksen, K. Balog . A. P. de Vries: *REL: An Entity Linker Standing on the Shoulders of Giants*. . *Proc. Int. ACM SIGIR Conf. on Research and Development in Information Retrieval*, . 2197–2200, Xi'an, China, 2020

REL: Radbound Entity Linker

Mention Detection

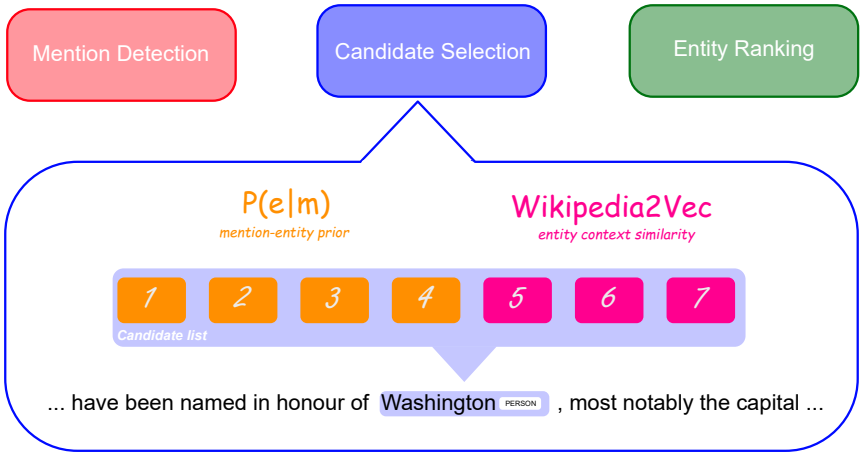
Candidate Selection

Entity Ranking



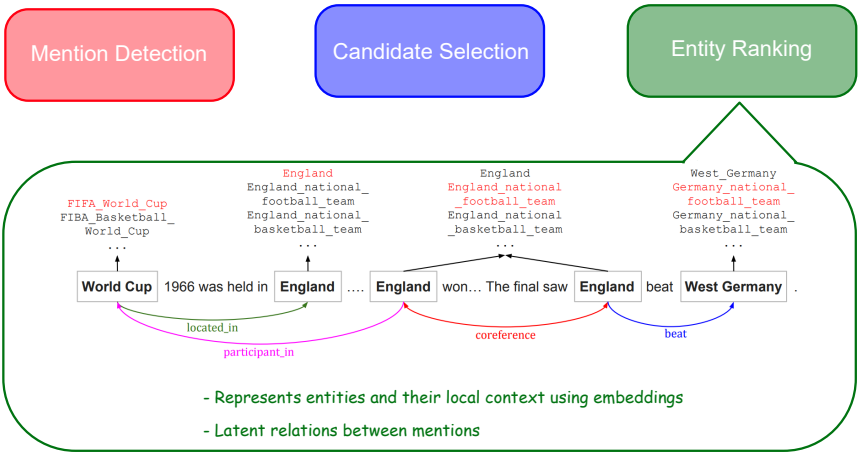
[1] A. Akbik, T. Bergmann, D. Blythe, K. Rasul, S. Schweter . R. Vollgraf: *FLAIR: An Easy-to-Use Framework for State-of-the-Art NLP.* . *Annual Conf. of the North American Chapter of the Association for Computational Linguistics*, . 54–59, Minneapolis, MN, USA, 2019

REL: Radbound Entity Linker



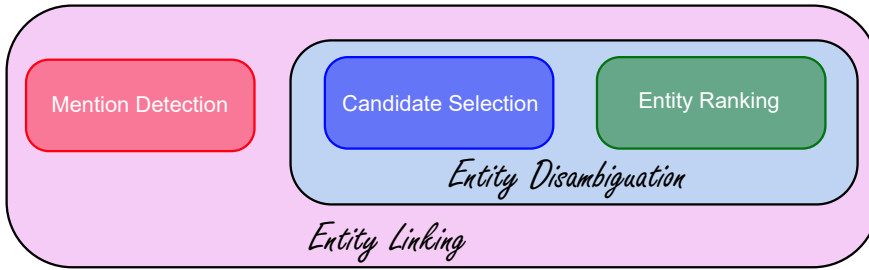
[1] O. Ganea . T. Hofmann: *Deep Joint Entity Disambiguation with Local Neural Attention*. . *Proc. Conf. on Empirical Methods in Natural Language Processing*, . 2619–2629, Copenhagen, Denmark, 2017

REL: Radbound Entity Linker



[1] P. Le . I. Titov: *Improving Entity Linking by Modeling Latent Relations between Mentions*. . *Annual Meeting of the Association for Computational Linguistics*, . 1595–1604, Melbourne, Australia, 2018

REL: Radbound Entity Linker



Evaluation metric

$$\textit{Precision} = \frac{\#\text{ of correctly detected and disambiguated mentions}}{\#\text{ of predicted mentions by model}} \tag{1}$$

$$\textit{Recall} = \frac{\#\text{ of correctly detected and disambiguated mentions}}{\#\text{ of mentions in gold standard}} \tag{2}$$

$$F1 = \frac{2 * \textit{Precision} * \textit{Recall}}{\textit{Precision} + \textit{Recall}} \tag{3}$$

InKB $\rightarrow$ only mentions that have an entity in KB or where NEL approach predicted an entity for

Strong annotation matching $\rightarrow$ mention has to be recognized completely and linked properly

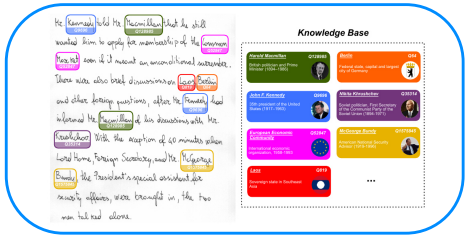
Results

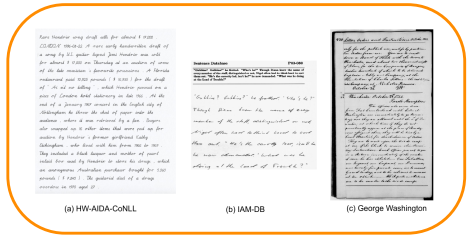
Table: Named Entity Disambiguation (NED) and Linking (NEL) performances measured in InKB micro F1 score based on strong annotation matching

		AIDA-B	MSNBC	WIKI	CWEB	IAM	GW
NED	*Annotation*	87.9	91.9	75.7	76.7	69.4	22.6
	HTR+D	74.8	80.1	63.1	60.1	54.9	19.4
	HTR	52.0	59.0	47.9	45.8	31.4	12.9
NEL	*Annotation*	79.1	75.4	40.1	47.8	49.8	15.2
	HTR+D	68.0	64.2	33.2	34.9	38.1	12.5
	HTR	56.7	55.7	30.6	33.8	27.5	7.1

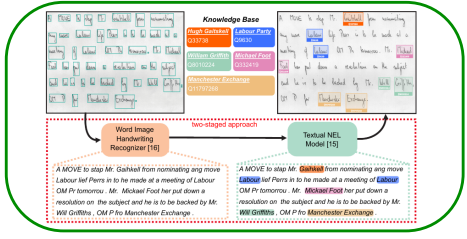
technische universität
dortmund

Conclusions





(a) HW-AIDA-CoNLL (b) IAM-DB (c) George Washington



MSNBC	WIKI	CWEB	IAM	GW
91.9	75.7	76.7	69.4	22.6
80.1	63.1	60.1	54.9	19.4
59.0	47.9	45.8	31.4	12.9
75.4	40.1	47.8	49.8	15.2
64.2	33.2	34.9	38.1	12.5
55.7	30.6	33.8	27.5	7.1