|  |  |  |
| --- | --- | --- |
| Data set | charcterstics | technique |
| NCBI-disease | consists of 22,137 sentences totally and is split into 14,987, 3,466 and 3,684 sen- tences for the training It includes annotations for 4 types of named entities: PERSON, LOCATION, ORGANIZATION and MISC | BERT, Bi-LSTM, CNN |
| Ontonotes v5 | made up of 1,745 K English, data.  full list of 18 categories | Transformer, LSTM,MLMET |
| **CoNLL 2003** | **It has 18 entities**  **It has 14,987 testing sets** | **knowledge distillation, GCN,**  **Bender** |

**BERT:**

1. Lee, Jinhyuk, Wonjin Yoon, Sungdong Kim, Donghyeon Kim, Sunkyu Kim, Chan Ho So, and Jaewoo Kang. "BioBERT: a pre-trained biomedical language representation model for biomedical text mining." *Bioinformatics* 36, no. 4 (2020): 1234-1240.

**Transformer:-**

1. Xu, Lu, Zhanming Jie, Wei Lu, and Lidong Bing. "Better Feature Integration for Named Entity Recognition." *arXiv preprint arXiv:2104.05316* (2021).

**knowledge distillation:-**

1. Zhou, Wenxuan, and Muhao Chen. "Learning from Noisy Labels for Entity-Centric Information Extraction." *arXiv preprint arXiv:2104.08656* (2021).

**Bi-LSTM:-**

1. Jang, Beakcheol, Myeonghwi Kim, Gaspard Harerimana, Sang-ug Kang, and Jong Wook Kim. "Bi-LSTM model to increase accuracy in text classification: Combining Word2vec CNN and attention mechanism." Applied Sciences 10, no. 17 (2020): 5841.

**CNN:-**

1. Zhu, Qile, Xiaolin Li, Ana Conesa, and Cécile Pereira. "GRAM-CNN: a deep learning approach with local context for named entity recognition in biomedical text." Bioinformatics 34, no. 9 (2018): 1547-1554.

**LSTM:-**

1. Cho, Minsoo, Jihwan Ha, Chihyun Park, and Sanghyun Park. "Combinatorial feature embedding based on CNN and LSTM for biomedical named entity recognition." Journal of biomedical informatics 103 (2020): 103381.

**MLMET:-**

|  |
| --- |
|  |
|  |

1. Dai, Hongliang, Yangqiu Song, and Haixun Wang. "Ultra-Fine Entity Typing with Weak Supervision from a Masked Language Model." arXiv preprint arXiv:2106.04098 (2021).

**GCN :-**

1. Seti, Xieraili, Aishan Wumaier, Turgen Yibulayin, Diliyaer Paerhati, Lulu Wang, and Alimu Saimaiti. "Named-entity recognition in sports field based on a character-level graph convolutional network." Information 11, no. 1 (2020): 30.

**Bender:-**

1. Li, Jing, Aixin Sun, Jianglei Han, and Chenliang Li. "A survey on deep learning for named entity recognition." IEEE Transactions on Knowledge and Data Engineering (2020).

<https://catalog.ldc.upenn.edu/docs/LDC2013T19/OntoNotes-Release-5.0.pdf>

https://towardsdatascience.com/deep-learning-for-ner-1-public-datasets-and-annotation-methods-8b1ad5e98caf#:~:text=OntoNotes%20Release%205.0%20is%20made,broadcast%20conversation%20and%20web%2Dblogs.