ANEA: Distant Supervision for Low-Resource Named Entity Recognition

Anonymous authors
Paper under double-blind review

Abstract

Distant supervision allows obtaining labeled training corpora for low-resource settings where only limited hand-annotated data exists. However, to be used effectively, the distant supervision must be easy to gather. In this work, we present ANEA, a tool to automatically annotate named entities in text based on entity lists. It spans the whole pipeline from obtaining the lists to analyzing the errors of the distant supervision. A tuning step allows the user to improve the automatic annotation with their linguistic insights without labelling or checking all tokens manually. In six low-resource scenarios, we show that the F1-score can be increased by on average 18 points through distantly supervised data obtained by ANEA.

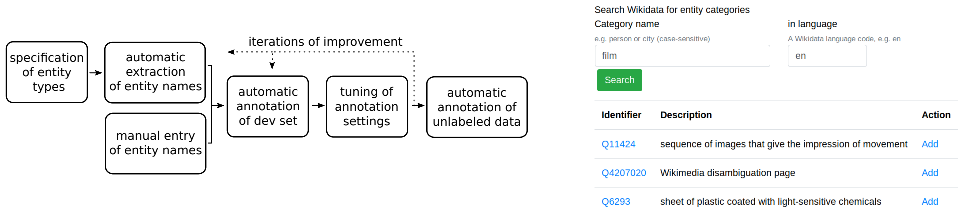
1 Introduction

Named Entity Recognition (NER) is a core NLP task necessary for various applications, from information retrieval to virtual assistants. While there exist some large, hand-annotated corpora like (Tjong Kim Sang & De Meulder, 2003) or (Weischedel et al., 2011), these are limited to a selected set of languages and domains. For many low-resource languages and domains, it is not possible to manually label every token of large corpora due to time and resource constraints. The absence of labeled data is prevalent for languages from developing countries. We see this as a significant factor limiting the development of NLP technologies in these regions with respect to the ongoing tendency towards data-driven models.

To overcome the lack of labeled data, weak or distant supervision methods have become popular, which automatically annotate unlabeled, raw text. Even in low-resource settings, unlabeled text is often available, and research has shown that automatically annotated labels can be a useful training resource in the absence of expensive, high-quality labels For NER, a widespread approach is to use lists, dictionaries or gazetteers of named entities (e.g. a list of person names or cities). Each word in the corpus is assigned the corresponding named entity label if it appears in this list of entities. Introduced by Mintz et al. (2009), this is still a popular technique and used e.g. by Peng et al. (2019), Adelani et al. (2020) and Lison et al. (2020). For an extensive list of recent works using distant supervision for low-resource NER, we refer to the recent survey by Hedderich et al. (2020).

While distant supervision performs very well on high-resource languages, it has been shown to be more difficult to leverage in real low-resource settings due to the lack of external information (Kann et al., 2020). Additionally, several difficulties arise when applying it in a practical way, such as obtaining these dictionaries (e.g. a list of city names in Yorùbá) or adapting the matching procedure to the specific language and domain (e.g. deciding for or against lemmatization and, thus, trading off recall and precision). Distant supervision can only be beneficial and save resources if it is easy to use and fast to deploy.

The ANEA tool we present provides the functionality to actually use distant supervision approaches in practice for many languages and named entity types while minimizing the amount of manual effort and labeling cost. A process is provided to automatically extract entity names from Wikidata, a free and open knowledge base. The information is used to annotate named entities for large amounts of unlabeled text automatically. The tool also supports the user in tuning the automatic annotation process. It enables language experts to efficiently include their knowledge without having to annotate many tokens manually. Both a library and a graphical user interface are provided to assist users of varying technical backgrounds and different use-cases. In an experimental study on


specification of entity types
automatic extraction of entity names
manual entry of entity names
iterations of improvement
automatic annotation of dev set
tuning of annotation settings
automatic annotation of unlabeled data

Search Wikidata for entity categories

Category name

e.g. person or city (case-sensitive)

film

in language

A Wikidata language code, e.g. en

en

Identifier	Description	Action
Q11424	sequence of images that give the impression of movement	Add
Q4207020	Wikimedia disambiguation page	Add
Q6293	sheet of plastic coated with light-sensitive chemicals	Add

(a) (b)

Figure 1: (a) Overall workflow of ANEA. (b) Interface to search for Wikidata categories from which to extract entity names.

six different scenarios, we show that ANEA outperforms two baselines in nearly all cases regarding the quality of the automatic annotation. When used to provide distantly supervised training data for a neural network model, it creates on average a boost of 18 F1 points with less than 30 minutes of manual interaction. The tool, further information and technical documentation and the additional model code and evaluation data will be made publicly available online[1].

2 Related Work

A variety of open-source tools exist to annotate text manually. While their focus is on the manual annotation of data, some support the user with certain degrees of automation. A token can be labeled automatically if it has been labeled before by the user in WebAnno (Yimam et al., 2014) and TALEN (Mayhew & Roth, 2018). In TALEN, a bilingual lexicon can be integrated but just to support annotators that do not speak the text's language. WebAnno and brat (Stenetorp et al., 2012) allow importing the annotations of external tools as suggestions for the user. The focus is, however, still on the user manually checking all tokens. Also, the annotator cannot use their insight to directly influence and improve the external tool like in the tuning process of ANEA.

In the area of information extraction, the tools by Gupta & Manning (2014), Li et al. (2015) and Dalvi et al. (2016) allow the user to create rules or patterns, e.g. "[Material] conducts [Energy]". They can, however, require a large amount of manual rule creation effort to obtain good coverage for NER. With Snorkel (Ratner et al., 2019), a user can define similar and more general labeling functions. Oiwa et al. (2017) presented a tool to create entity lists manually. These lists could be imported into ANEA. NER is closely related to entity linking. Zhang et al. (2018) presented a system to link entities in many languages automatically but focus on disaster monitoring and, therefore, only consider persons, geopolitical entities, organizations, and locations.

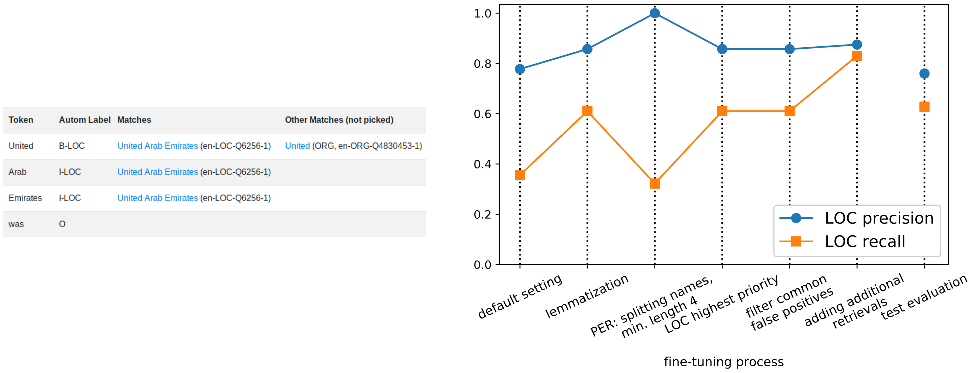
3 Workflow

The workflow is visualized in Figure 1a and we provide an online video that shows an exemplary walkthrough[2]. The process is split into four parts:

Extraction: The user starts by searching for the category names of the entity types that should be extracted (e.g. *person* or *film*). The tool will then automatically extract the names of all the corresponding entities (e.g. for *person*: "Alan Turing", "Edward Sapir", ...). As the source for the extractions, we use a dump of Wikidata. It is a free and open knowledge base that is created both by manual edits and automatic processes. At the time of writing, it contains over 90 million items. For most items, the names are available in multiple languages (e.g. 32k person names for Yorùbá or 26k movie names for Spanish). The user searches for and specifies the entity types they want to extract and which language should be used for the names (Figure 1b). The tool will then extract all items that have the "is an instance of" property of the given entity types. The results are the lists of entity names. Additionally, the user can also provide existing lists of entity names in case of a very specific domain.

[1] Anonymized code upload for submission `https://www.dropbox.com/s/lk9hdd9lxac4hiy/ANEA.zip`

[2] Anonymized video upload for submission `https://www.dropbox.com/s/uf8ztucooexwnm0/ANEA.mp4`



Token	Autom Label	Matches	Other Matches (not picked)
United	B-LOC	United Arab Emirates (en-LOC-Q6256-1)	United (ORG, en-ORG-Q4830453-1)
Arab	I-LOC	United Arab Emirates (en-LOC-Q6256-1)	
Emirates	I-LOC	United Arab Emirates (en-LOC-Q6256-1)	
was	O		

(a)

(b)

Figure 2: (a) Interface to manually inspect the automatic labeling. (b) Development of precision and recall during the tuning process on the Estonian data. On the x-axis, the setting changes over time are reported.

Automatic Annotation: The automatic annotation is performed by checking each word against the list of extracted entities. A word (or token) is assigned the label of the entity name it matches. If matches of several entity names overlap, the longest match is used. I.e. for the string “United Arab Emirates” the entity name of the country is preferred over the substring “United” (the airline) if both are in lists of entities.

Evaluation: If a small set of labeled data exists, it can be used to evaluate the automatic annotation. The tool can calculate precision, recall and F1-score directly. It also reports the tokens that were most often labeled incorrectly or not labeled. For a more in-depth analysis, for each token, one can check which label was assigned, which alternative labels could have been assigned and to which entities they correspond. This allows a user to easily understand issues of the automatic annotation (Figure 2a). Specific labels can also be changed manually.

Tuning: ANEA provides multiple options with which the automatic annotation can be improved. Guided by the evaluation from the previous step, this allows the user to easily insert language expertise into the annotation process and prevent common mistakes while still avoiding to annotate or post-edit many tokens manually. The options include lemmatization, filtering common false positives, stopword removal, adding alias names (like ”ICLR” for the ”International Conference on Learning Representations”), splitting entity names, removing diacritics, requiring a minimum length for the entities, prioritization of lists for resolution of conflicts or fuzzy matching of entities.

The effects of such a tuning process are visualized in Figure 2b for an Estonian dataset and the *location* label. Adding lemmatization in tuning-step 1 increases recall due to the language’s rich morphological structure that can hinder the matching. In step 3, location entities are given a higher priority if they conflict with person entities on the same token. In the last tuning-step, another gain can be obtained by extracting additional entity lists for Estonian locations based on the evaluation feedback. After the (optional) tuning process, unlabeled text can be automatically annotated for use as distant supervision.

4 Experimental Evaluation

4.1 Datasets

We selected a variety of datasets that reflect different languages and entity granularities. The first 1500 tokens of each dataset are used as labeled training instances. Garrette & Baldridge (2013) reported this as the number of tokens that can be annotated within two hours for a low-resource POS task. We think that this is a reasonable amount of labeled data that one can expect even in a low-resource setting, and it is also necessary for training the baselines we compare to. For **English (En)**, the CoNLL03 dataset is probably the most popular NER dataset. It was created for the CoNLL-2003 shared task (Tjong Kim Sang & De Meulder, 2003). To obtain a more specialized domain, we manually annotated the *location* labels from the CoNLL03 dataset with more specific labels. For **Spanish (Es)**, we manually annotated news articles with the label *movie* to resemble a Latin-American setting where e.g. a start-up requires a fine-grained and less common label. For **Yorùbá (Yo)**, a language spoken predominantly in West Africa, we evaluate on the dataset by Alabi et al. (2020). We also evaluate on two European low-resource languages, namely **Estonian (Et)** (Tkachenko et al., 2013) and **West Frisian (Fy)** (Pan et al., 2017). All results are reported on held-out test sets. The manually labeled data created for this evaluation will be made publicly available.

4.2 Machine Learning Models

We evaluate against two baselines that should, like ANEA, be easy and quick to use, do not require extensive development of hand-engineered features and do not have large hardware requirements. The Stanford NER tagger (Finkel et al., 2005) is a popular tool based on Conditional-Random-Fields (**CRF**) which we use in their suggested configuration[3]. For the second baseline, a neural network (**NN**), we performed preliminary experiments on held-out, English data in a low-resource setting and chose a combination of a bidirectional Gated Recurrent Unit (Cho et al., 2014) and a ReLU with Dropout (Srivastava et al., 2014) between the layers. To easily apply the model to many different languages, we used pretrained fastText embeddings (Grave et al., 2018) which are available in 157 languages. Model details are given in the code. In the high-resource setting on the full CoNLL03 dataset (>250k labeled tokens), both baselines achieve an F1-score of 87.

4.3 Experimental Setup

Experiment A: Here, the quality of the automatic annotation is evaluated. The CRF is trained on the 1500 labeled training tokens of each dataset. Similarly, for the neural network, the first 1000 tokens are used for the training. The remaining 500 tokens are held-out as the development set to select the best performing epoch and avoid overfitting. For ANEA, we report the scores with and without the tuning phase. *ANEA No Tuning* just uses the default settings without any labeled supervision and no manual interaction. For *ANEA + Tuning*, the 1500 labeled training token are used for the manual tuning. It was limited to no more than 10 manual steps and 30 minutes of user interaction per dataset.

Experiment B: For evaluating the effect of the distant supervision, unlabeled tokens are automatically annotated by the CRF, the NN and ANEA with Tuning. The NN model is then retrained on both the manually labeled and the distantly-supervised instances. 200k tokens from each of the datasets are used as unlabeled data. For Spanish, West Frisian and Yorùbá, ca. 15k and 70k and 18k tokens are used, respectively, due to the smaller dataset sizes. These texts are disjoint of the labeled training and test data.

4.3.1 Results

The results of Experiment A are given in Table 1a. The CRF approach can provide a high precision but often has a very low recall due to the limited amount of training data. The NN can leverage the pre-training of the embeddings on large amounts of unlabeled text. However, the training data seems not enough to reach a competitive performance. Our tool struggles most with organizations as these are stored as several different entity types in Wikidata. Another issue is the existence of false positives of words that have other meanings beyond entity names, e.g. the Turkish city “Of”. Nevertheless, reasonable results are obtained even if the amount of labeled tokens is too low for the baselines to learn anything meaningful (cf. *En CONTINENT* or *Et ORG*). Even without any labeled data, we are often able to reach competitive performance. Using the tuning process is helpful to boost the performance further. The possibility for the user to trade-off precision and recall can be seen in several cases (e.g. *En LOC* or *Et PER*). Overall, ANEA outperforms the other baselines in all metrics in a majority of the settings. It achieves the best F1-score in all but one case.

The higher quality of the automatic annotation is also reflected in Experiment B (Table 1b). For 14 out of 16 evaluated entity types, the distant supervision provided by ANEA achieves the largest improvements. On average, it increases the classifier’s performance by 18 points F1-score.

[3] `https://nlp.stanford.edu/software/crf-faq.html#a`

	CRF			NN			ANEA No Tuning			ANEA + Tuning		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
En PER	**75**	14	23	54	40	46	36	**51**	42	67	49	**57**
En LOC	66	22	33	54	52	52	**70**	45	55	56	**74**	**64**
En ORG	**24**	08	12	23	**13**	**16**	17	07	10	21	09	13
En CITY	**100**	14	25	27	43	33	16	30	21	29	**51**	**37**
En COUN.	**94**	05	10	63	51	56	93	80	86	84	**90**	**87**
En CONTI.	00	00	00	00	00	00	**75**	**94**	**83**	**75**	**94**	**83**
Es MOVIE	**75**	02	05	08	07	08	32	35	33	40	**40**	**40**
Et PER	66	24	35	61	30	40	**75**	17	27	41	**51**	**45**
Et LOC	59	27	37	44	25	32	71	36	48	**76**	**63**	**69**
Et ORG	00	00	00	17	09	12	75	12	21	**81**	**17**	**29**
Fy PER	07	06	07	04	03	04	**55**	**42**	**48**	**55**	**42**	**48**
Fy LOC	32	55	41	33	42	37	**68**	24	37	61	**34**	**43**
Fy ORG	00	00	00	00	00	00	89	07	13	**90**	**08**	**14**
Yo PER	33	05	10	15	22	18	11	13	12	**49**	**43**	**46**
Yo LOC	**100**	07	12	48	27	35	64	72	68	65	**74**	**69**
Yo ORG	00	00	00	07	08	08	16	28	20	**46**	**52**	**49**

(a)

	NN + Distant Supervision by ...		
	CRF	NN	ANEA
En PER	-35	+5	**+15**
En LOC	-20	+1	**+13**
En ORG	-6	**0**	-5
En CITY	-13	+1	**+6**
En COUN.	-45	-6	**+30**
En CONTI.	0	0	**+88**
Es Movie	-7	+2	**+14**
Et PER	-7	-7	**+14**
Et LOC	+10	-1	**+39**
Et ORG	-2	0	**+17**
Fy PER	+1	0	**+26**
Fy LOC	**+4**	+1	**+4**
Fy ORG	+1	+1	**+7**
Yo PER	-4	**+6**	-5
Yo LOC	-25	+4	**+5**
Yo ORG	-1	+1	**+20**

(b)

Table 1: Results of Experiment A (a) and Experiment B (b) on the test data. We report precision/recall/F1-score in percentage (higher is better).

5 Technical Aspects

The tool consists of both a library for the core functionalities as well as a graphical user interface. The user can control the interface in the browser with the back end running on the local system. Alternatively, the back end can run on a different, more powerful machine and is then accessed remotely. All the code is published as open-source under the Apache 2 license, and we welcome contributions from other authors. The tool is implemented in Python 3 using Flask[4] for the web-server's back end and Bootstrap 4[5] for the front end. To overcome the rate limitations of the Wikidata Web API, a database dump of Wikidata is used. To reduce hardware requirements, care was taken during the implementation to limit the memory footprint.

The user can upload text files or insert them directly into a text field. For labeled data, the CoNLL column format is supported. Annotated text can be downloaded in the same format. Tokenization and lemmatization are provided for a variety of languages via the SpaCy (Honnibal et al., 2020) and EstNLTK (Laur et al., 2020). For other languages, the text can be preprocessed with an external system before inputting it, or the external tool can be easily integrated into ANEA. Stopword lists for 58 languages are included.

6 Conclusion

We presented an open-source tool to obtain large amounts of distantly supervised training data for NER in a quick way and with few manual efforts and costs. While the annotation itself is automatic, the user can tune it to add their expertise. To support users of varying technical backgrounds, both a library and a graphical user interface are provided. The experiments showed its usefulness in six different language and domain settings.

In the future, we aim to add techniques from active learning to improve the efficient leverage of expert insights further. Also, recent works have shown that the performance gains through distant supervision can be further boosted by handling errors in the automatic annotation via label noise modeling (Lange et al., 2019) or filtering (Le & Titov, 2019). We see the integration of these approaches as an additional avenue for interesting future work.

[4] http://flask.pocoo.org

[5] https://getbootstrap.com

REFERENCES

David Ifeoluwa Adelani, Michael A. Hedderich, Dawei Zhu, Esther van den Berg, and Dietrich Klakow. Distant supervision and noisy label learning for low resource named entity recognition: A study on hausa and yorùbá, 2020.

Jesujoba O Alabi, Kwabena Amponsah-Kaakyire, David I Adelani, and Cristina España-Bonet. Massive vs. curated word embeddings for low-resourced languages. the case of yorùbá and Twi. In *Proc. of LREC 2020*, 2020.

Kyunghyun Cho, Bart van Merrienboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using rnn encoder–decoder for statistical machine translation. In *Proc. of EMNLP 2014*, 2014. doi: 10.3115/v1/D14-1179. URL `http://aclweb.org/anthology/D14-1179`.

Bhavana Dalvi, Sumithra Bhakthavatsalam, Chris Clark, Peter Clark, Oren Etzioni, Anthony Fader, and Dirk Groeneveld. IKE - an interactive tool for knowledge extraction. In *Proc. of AKBC 2016*, 2016. doi: 10.18653/v1/W16-1303.

Jenny Rose Finkel, Trond Grenager, and Christopher Manning. Incorporating non-local information into information extraction systems by gibbs sampling. In *Proc. of ACL 2005*, 2005. URL `http://aclweb.org/anthology/P05-1045`.

Dan Garrette and Jason Baldridge. Learning a part-of-speech tagger from two hours of annotation. In *Proc. of NAACL 2013*, 2013. URL `https://www.aclweb.org/anthology/N13-1014`.

Edouard Grave, Piotr Bojanowski, Prakhar Gupta, Armand Joulin, and Tomas Mikolov. Learning word vectors for 157 languages. In *Proc. of LREC 2018*, 2018. URL `https://www.aclweb.org/anthology/L18-1550`.

Sonal Gupta and Christopher Manning. SPIED: Stanford pattern based information extraction and diagnostics. In *Proc. of the Workshop on Interactive Language Learning, Visualization, and Interfaces*, 2014. doi: 10.3115/v1/W14-3106.

Michael A. Hedderich, Lukas Lange, Heike Adel, Jannik Strötgen, and Dietrich Klakow. A survey on recent approaches for natural language processing in low-resource scenarios, 2020.

Matthew Honnibal, Ines Montani, Sofie Van Landeghem, and Adriane Boyd. spaCy: Industrial-strength Natural Language Processing in Python, 2020. URL `https://doi.org/10.5281/zenodo.1212303`.

Katharina Kann, Ophélie Lacroix, and Anders Søgaard. Weakly supervised POS taggers perform poorly on *Truly* low-resource languages. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020*, pp. 8066–8073. AAAI Press, 2020. URL `https://aaai.org/ojs/index.php/AAAI/article/view/6317`.

Lukas Lange, Michael A. Hedderich, and Dietrich Klakow. Feature-dependent confusion matrices for low-resource ner labeling with noisy labels. In *Proc. of EMNLP 2019*, 2019.

Sven Laur, Siim Orasmaa, Dage Särg, and Paul Tammo. Estnltk 1.6: Remastered estonian nlp pipeline. In *Proceedings of The 12th Language Resources and Evaluation Conference*, pp. 7154–7162, Marseille, France, May 2020. European Language Resources Association. URL `https://www.aclweb.org/anthology/2020.lrec-1.884`.

Phong Le and Ivan Titov. Distant learning for entity linking with automatic noise detection. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pp. 4081–4090, Florence, Italy, July 2019. Association for Computational Linguistics. doi: 10.18653/v1/P19-1400. URL `https://www.aclweb.org/anthology/P19-1400`.

Yunyao Li, Elmer Kim, Marc A. Touchette, Ramiya Venkatachalam, and Hao Wang. Vinery: A visual ide for information extraction. *Proc. of the VLDB Endowment*, 8(12), 2015. doi: 10.14778/2824032.2824108.

Pierre Lison, Jeremy Barnes, Aliaksandr Hubin, and Samia Touileb. Named entity recognition without labelled data: A weak supervision approach. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 1518–1533, Online, July 2020. Association for Computational Linguistics. URL `https://www.aclweb.org/anthology/2020.acl-main.139`.

Stephen Mayhew and Dan Roth. Talen: Tool for annotation of low-resource entities. In *Proc. of ACL 2018: System Demonstrations*, 2018. URL `http://aclweb.org/anthology/P18-4014`.

Mike Mintz, Steven Bills, Rion Snow, and Daniel Jurafsky. Distant supervision for relation extraction without labeled data. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, pp. 1003–1011, Suntec, Singapore, August 2009. Association for Computational Linguistics. URL `https://www.aclweb.org/anthology/P09-1113`.

Hidekazu Oiwa, Yoshihiko Suhara, Jiyu Komiya, and Andrei Lopatenko. A lightweight front-end tool for interactive entity population. *CoRR*, abs/1708.00481, 2017. URL `http://arxiv.org/abs/1708.00481`.

Xiaoman Pan, Boliang Zhang, Jonathan May, Joel Nothman, Kevin Knight, and Heng Ji. Cross-lingual name tagging and linking for 282 languages. In *Proc. of ACL 2017*, 2017. doi: 10.18653/v1/P17-1178.

Minlong Peng, Xiaoyu Xing, Qi Zhang, Jinlan Fu, and Xuanjing Huang. Distantly supervised named entity recognition using positive-unlabeled learning. In *Proc. of ACL 2019*, 2019. doi: 10.18653/v1/P19-1231. URL `https://www.aclweb.org/anthology/P19-1231`.

Alexander Ratner, Stephen H. Bach, Henry Ehrenberg, Jason Fries, Sen Wu, and Christopher Ré. Snorkel: rapid training data creation with weak supervision. *The VLDB Journal*, Jul 2019. ISSN 0949-877X. doi: 10.1007/s00778-019-00552-1. URL `https://doi.org/10.1007/s00778-019-00552-1`.

Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.*, 15(1), 2014.

Pontus Stenetorp, Sampo Pyysalo, Goran Topić, Tomoko Ohta, Sophia Ananiadou, and Jun'ichi Tsujii. brat: a web-based tool for nlp-assisted text annotation. In *Proc. of the Demonstrations at EACL 2012*, 2012. URL `http://aclweb.org/anthology/E12-2021`.

Erik F Tjong Kim Sang and Fien De Meulder. Introduction to the conll-2003 shared task: Language-independent named entity recognition. In *Proc. of the Seventh Conference on Natural Language Learning*, 2003. URL `https://www.aclweb.org/anthology/W03-0419`.

Alexander Tkachenko, Timo Petmanson, and Sven Laur. Named entity recognition in estonian. In *Proc. of the 4th Biennial International Workshop on Balto-Slavic Natural Language Processing*, 2013.

Ralph Weischedel, Sameer Pradhan, Lance Ramshaw, Martha Palmer, Nianwen Xue, Mitchell Marcus, Ann Taylor, Craig Greenberg, Eduard Hovy, Robert Belvin, et al. Ontonotes release 4.0. *LDC2011T03*, 2011.

Seid Muhie Yimam, Chris Biemann, Richard Eckart de Castilho, and Iryna Gurevych. Automatic annotation suggestions and custom annotation layers in webanno. In *Proc. of ACL 2014: System Demonstrations*, 2014. doi: 10.3115/v1/P14-5016.

Boliang Zhang, Ying Lin, Xiaoman Pan, Di Lu, Jonathan May, Kevin Knight, and Heng Ji. Elisa-edl: A cross-lingual entity extraction, linking and localization system. In *Proc. of NAACL-HLT 2018: Demonstrations*, 2018. doi: 10.18653/v1/N18-5009. URL `http://aclweb.org/anthology/N18-5009`.