An Overview of Named Entity Recognition

Peng Sun*†, Xuezhen Yang*, Xiaobing Zhao*† and Zhijuan Wang*†‡
*School of Information Engineering
Minzu University of China, Beijing, China 100081
†Minority Languages Branch, National Language Resource and Monitoring Research Center
‡Email: wangzj_muc@126.com

***Abstract*—Named Entity Recognition (NER) is essential for some Natural Language Processing (NLP) tasks. Previous researchers gave a survey of NER in statistical machine learning era, however, research on NER has already changed a lot in recent decade. On the one hand, more and more NER systems adopt deep learning, transfer learning, knowledge base and other methods. On the other hand, multilingual and low resource languages NER researches increase rapidly. To reflect these changes, we here give an overview of NER based on 162 papers of NLP related conferences from 1996 to 2017. In this survey, we discuss two main aspects of NER research - target languages and technical approaches with statistical analysis. Finally, we summarize some conclusions and explore potential future issues in NER research.**

Keywords-named entity recognition; deep learning; transfer learning; multilingual NER

I. Introduction

Named Entity Recognition is important for some Natural Language Processing tasks such as Information Extraction (IE) and Question Answering (QA). Its goal is to recognize and classify names of person (PER), location (LOC), organization (ORG), and numeric expressions including date, currency and percentage. Research on NER has a history of more than 20 years since the Sixth Message Understanding Conference (MUC-6) [1] that concentrated on IE tasks where organized information of organization activities and security affairs is recognized from raw content. Early target languages of NER are several rich resource languages (e.g., English, Chinese). Now more and more low resource languages (e.g., Indian languages) attract researchers' attention. The approaches constructing NER systems also have greatly changed. Although rule-based methods still been adopted, much researchers focus on statistical machine learning. The earlier is machine learning (e.g., Conditional Random Fields, Hidden Markov Model, etc.), followed by deep learning (e.g., Recurrent Neural Networks) and other methods (e.g., transfer learning and knowledge base). To reflect the latest changes of research on NER, we retrieve well-known NLP conference website - ACL Anthology[1] with keywords "named entity" and select 162 papers to present an overview of research on NER.

These publications cover more than 200 target languages, spanning from 1996 to 2017, so we believe that they can provide a comprehensive perspective for both breadth and depth of NER related work. The first part focuses on analyzing target languages of NER in different dimensions. The second section lists and explains the algorithms proposed for coping with NER tasks. The third section combines the definition of approaches and the way of previous classification [2], we induce four categories as follows: rule-based, traditional machine learning, deep learning, and other methods (mainly include knowledge base and transfer learning). Finally, we present our conclusions and end the article with a discussion of future work.

[1] https://aclanthology.coli.uni-saarland.de/

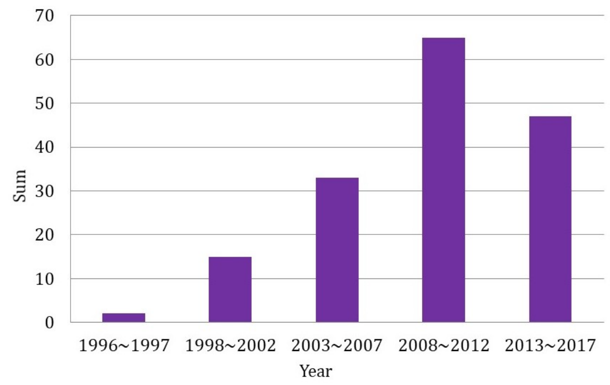

Figure 1. Number of NER papers published from 1996 to 2017.

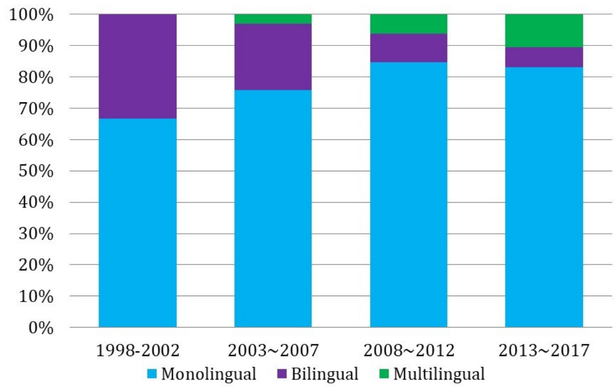

Figure 2. Proportion of monolingual, bilingual, multilingual NER papers for every stage (1998 to 2017).

II. Target Languages

To archive better performance, it is very necessary for NER systems to understand the characteristics of target languages. Sometimes, existing research on one or some

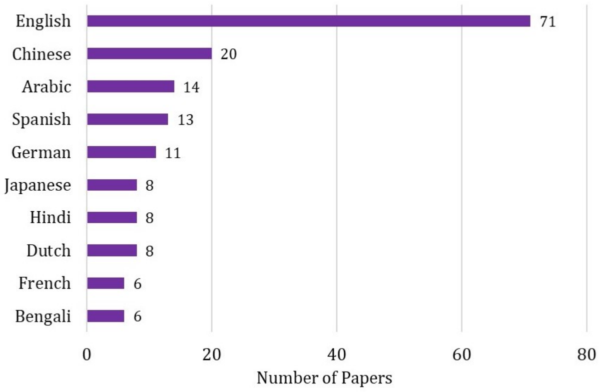

Figure 3. Top 10 target languages that receive most attention.

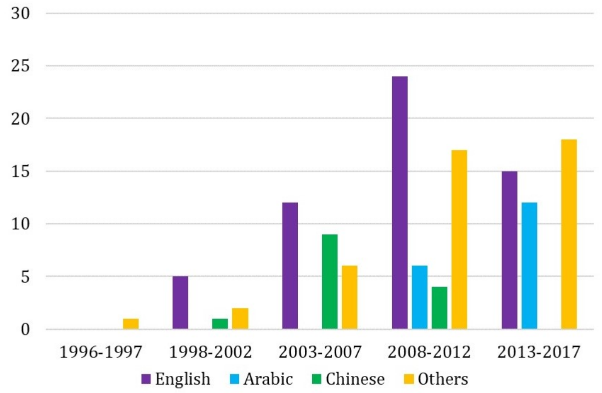

Figure 4. Number of English, Arabic, Chinese, and other languages NER Papers.

languages can bring inspiration to new target languages, including approaches and data. Target languages often are decisive for NER's strategies. This chapter focuses on target languages and illustrates multiple aspects statistics features of these NER articles.

Fig. 1 lists the number distribution of all 162 papers for every stage that has five years except for the last stage. In general, the number of papers on NER increases gradually. There are many NER systems simultaneously address more than one target language. Fig. 2 illustrates the proportion of monolingual, bilingual, multilingual NER papers at every stage. Obviously, we can learn that multilingual NER research is getting popular. Fig. 3 illustrates top 10 target languages receiving most attention, and we surprisingly know that some languages (e.g., Arabic, probably is regarded as low resource languages) NER studies have already made such great progress, what's more, keep increasing. Even if English still plays an absolute leadership role in NER research, research on NER of other languages is developing rapidly as Fig. 4 shows.

A. Monolingual Named Entity Recognition

Monolingual named entity recognition refers to NER system that deals with single target language. There are 130 monolingual NER papers covering 29 languages. Fig. 3 shows that English has always been the hot spot of research NER papers, followed by Arabic and Chinese, and other languages such as Hindi. Other low resource languages are gradually been valued. Monolingual named entity recognition still plays a leading role on research.

Table I
LANGUAGE PAIR IN BILINGUAL NER

Language Pair	Number of Papers	Percentage
English-German	6	28.4%
Dutch-Spanish	5	23.8%
English-Chinese	2	9.5%
Portuguese-Spanish	2	9.5%
Bulgarian-Korean	1	4.8%
Chinese-Spanish	1	4.8%
English-Hindi	1	4.8%
English-Italian	1	4.8%
English-Hungarian	1	4.8%
Greek-French	1	4.8%
Total	21	100%

Table II
TARGET LANGUAGES NUMBER IN MULTILINGUAL NER

Target Languages Number	Number of Papers	Percentage
Three Languages	4	36.3%
Four Languages	1	9.1%
Five Languages	1	9.1%
Six Languages	3	27.3%
10 Languages	1	9.1%
282 Languages	1	9.1%
Total	11	100%

B. Bilingual Named Entity Recognition

Bilingual named entity recognition refers to NER system that addresses two target languages. Table 1 describes the language pairs involved in bilingual NER papers. A total of eight language pairs are involved. Both English-German and English-Dutch language pairs reach 30% in all bilingual NER papers. Ten papers covers four pairs of languages (English-German, English-Chinese, English-Italian, and English-Hindi), accounting for 50% of bilingual NER papers.

C. Multilingual Named Entity Recognition

Multilingual named entity recognition refers to NER system that deals with more than two target languages. Palmer and Day [3] firstly studied the multilingual NER. Table 2 shows target language groups and their paper numbers in multilingual NER. Triple-language NER has most papers, accounting for 36%, followed by papers in six-language, accounting for approximately 28%. Pan et al. [4] proposed a cross-lingual named entity labeling and linking structure that recognizes 282 languages supported by Wikipedia. Table 3 illustrates the regional distribution of target languages involved in multilingual NER, languages originated from European and Asian languages are more active.

Table III
REGIONAL DISTRIBUTION OF LANGUAGES IN MULTILINGUAL NER

Area	Country	Language	Number of Papers
-	Many Countries	English	2
Europe	France	French	2
	Portugal	Portuguese	3
	Spain	Spanish	4
	Poland	Polish	1
	Germany	German	4
	Netherlands	Dutch	2
	Russia	Russian	1
	Ukraine	Ukrainian	1
Asia	India	Hindi	5
		Telugu	3
		Oriya	1
	Bengal, India	Bengali	5
	Pakistan, India	Urdu	1
	China	Chinese	1
	Japan	Japanese	2
	North/South Korea	Korean	1

Table IV
MATRIX OF METHODS AND LANGUAGES

	Monolingual	Bilingual	Multilingual
Rule-based	23 (18%)	3 (14%)	1 (9%)
Machine Learning	85 (65%)	13 (62%)	5 (45%)
Deep Learning	12 (9%)	2 (10%)	2 (18%)
Others	10 (8%)	3 (14%)	3 (28%)
Total	130 (100%)	21 (100%)	11 (100%)

III. APPROACHES OF NAMED ENTITY RECOGNITION

The approaches of NER strongly rely on the availability of labeled language resource. Most of theories have been proposed for many years, but they are induced to solve NER tasks in recent years. Early researches were mainly based on handcrafted rules and small-scale corpora, recent approaches include machine learning and deep learning. On the one hand, the approaches combining rules and machine learning or machine learning and deep learning have also been adopted for rich resource languages. On the other hand, many low resource languages still use rules and traditional machine learning to cope with NER tasks. As we have proposed at section one, there are four types of methods for building NER systems. In this chapter, we firstly give a quantitative overview of all types. Then, each type of method and its usage will be exampled from references.

We have calculated the proportion of rule-based, traditional machine learning, deep learning, and others for total papers, Fig. 5 illustrates that top 2 approaches are traditional machine learning and rule-based approaches. Table 4, the matrix of methods and languages, verifies the above conclusion. When we extend the time-line and subdivide it into every period (Fig. 6 distinguishes TL+KB from other methods, in fact, the latter includes the former in other sections of this paper), we learn the changes that each kind of method plays different role at every stage. In general, rule-based and traditional machine learning for NER gradually decreases, deep learning and others increase.

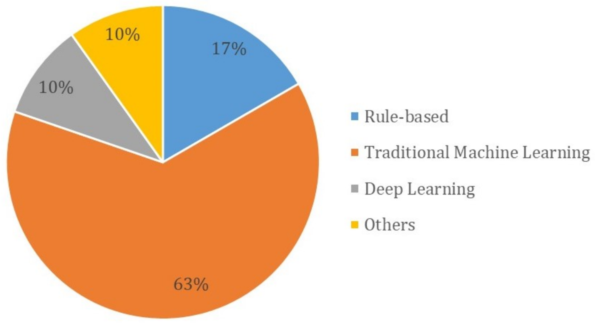

Figure 5. Proportion of Different NER Methods Papers.

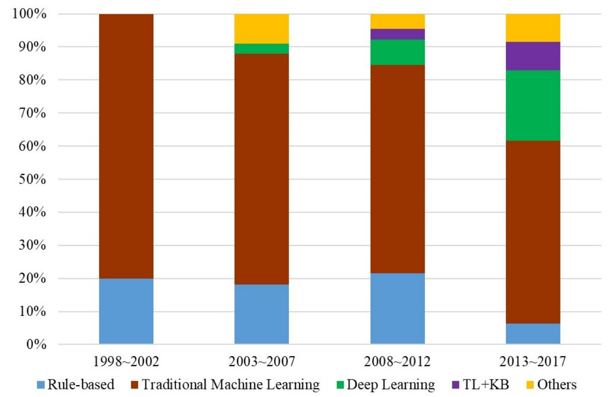

Figure 6. Distribution of Different Methods for Every Stage. TL+KB represents Transfer Learning and Knowledge Base.

A. Rule-based

Rule-based approaches are intuitionistic to recognize named entities with the help of language-specific features. There are three ways using rules to deal with NER tasks: rules (lexical rules, contextual rules, morphological lexicon), combination of rules and dictionaries (Gazetteers), combination of rules and feature selection algorithms. Consider some obvious morphology of named entities, regular expression can match most data, time, money and percent entities.

Kashif Riaz [5] presented a rule-based Urdu NER algorithm that outperforms the model with statistical learning. Demiros et al. [6] presented a rule-based NER system combining Gazetteers in Greek texts. Tomoya Iwakura [7] proposed a boosting Japanese NER method with rules acquired from unlabeled data.

Although rule-based methods are no longer mainstream approaches, they are still been adopted in NER tasks. For example, Aurore de Amaral [8] presented a tool which detects proper names, locations and dates from French texts by manually written linguistic rules. Whats more, rule-based methods still appear in some low resource languages NER systems [9], [10], [11].

B. Machine Learning (Nadeau and Sekine, 2007)

Approaches to machine learning are based on large-scale labeled data set. With the development of NER research, more and more researchers are attracted by systems built with less or no labeled corpus. General machine learning consists of supervised learning, semi-supervised learning and unsupervised learning. In recent years, mixed learning approaches combing rules and machine learning or multiply stacked machine learning are proposed.

1) Supervised Learning: Learning methods learn language knowledge by labeled training data and predict unlabeled test data. Supervised learning is one of dominant techniques for addressing NER problems, includes Conditional Random Fields (CRF) [12], Decision Trees [13], Hidden Markov Models (HMM) [14], Maximum Entropy Models (ME) [15], Support Vector Machines (SVM) [16], Boosting (Ada-Boost) [17]. We also notice that some systems are used as generalized boosting methods. For example, Wu et al. [18] investigated a stacking and voting method for jointing some effective classifiers such as boosting, SVM and Transformation-based Learning (TBL) on NER task, and demonstrated several effective methods culminating in single model outperforms standard baseline respectively.

2) Semi-supervised Learning: Semi-supervised learning leverages unlabeled data to moderate the impact of inadequate labeled data on system accuracy by learning to automatically make golden or silver standard data from unlabeled data. Typical supervised learning methods include CRF, ME, and SVM also are been used to construct semi-supervised learning NER system. Liao and Veeramachaneni [19], Kucuk and Steinberger [20] presented semi-supervised learning algorithm and system with CRF on Temil and Turkish. Nadeau and Sekine [2] introduced an approach called bootstrap that was adopted by Brooke et al. [21]. Althobaiti et al. [22] trained two NER classifiers, one with semi-supervised approach, another with distant learning technique, and finally combined both with a combination of classifier named Bayesian Classifier Combination. In general, semi-supervised learning methods seldom cope with NER task.

3) Unsupervised Learning: One of approaches to unsupervised learning is clustering-based method that coreference information plays an important role in NER task. Elsner et al. [23] described an entire unsupervised generative model with features of named entities conference information and syntactic context. Bhagavatula et al. [24] detailed a original clustering and co-occurrence based approach to build the mapping between English named entities and their equivalent representations from different languages that are detected in a language-independent way. Patil et al. [25] presented an enhanced version of bootstrapping that uses a new variant of Multi-word Expression Distance to quantify the proximity of a candidate phrase with a given NE type. Brooke et al. [21] presented a NER system that bootstraps a model from term clusters and uses several instances of the identical name in a text.

C. Deep Learning

Named Entity Recognition has needed large numbers of language knowledge in the way of lexicons and feature engineer to get higher performance. Deep learning is a branch of machine learning based on neural networks which have shown the effectiveness for NER tasks with automatically detecting word and character level features. Recently, the models based on neural networks mainly are used to solve English and Chinese NER tasks, often need large amount training data. Deep Learning widely becomes popular for NER since 2010. Recurrent Neural Networks (RNN) could capture context of sentence and is especially good at sequence task. Long Short-Term Memory (LSTM), an improved method of RNN, is adopted by more than 40% of papers with deep learning for NER.

1) LSTM: Long Short-Term Memory which could overcome long-distance dependencies does well in many NLP tasks. Chiu et al. [26] presented a hybrid bidirectional LSTM and CNN to reduce the need of feature engineering. James et al. [27] proposed a LSTM to perform two passes on each sentence. Guillaume et al. [28] designed a deep learning model based on bidirectional LSTMs and CRFs, this system obtained great performance without any language-specific knowledge. A model based on bidirectional LSTM can introduce a large number of every word of context on both sides and eliminate the issue of limited context applied to most feedforward models (Chiu et al. [26]).

2) DBN: Chen et al. [29] used a Deep Belief Nets (DBN) which can elaborately train by self for discovering complicated feature combination of Chinese named entity.

3) FFNN: Xu et al. [30] proposed a new local detection method instead of regrading NER as sequence labeling task and used a local feedforward neural network (FFNN) to predict named entity label.

4) CharWNN: Santos et al. [31] proposed a model, based on CharWNN model that units character and word level embedding to deal with sequential classification, does not rely on handcrafted features and outperforms the state-of-the-art system.

D. Other Methods

1) Knowledge Base: Knowledge Base (KB) based methods, usually learn and link external open source knowledge, are able to cope with multilingual NER task. Common KBs include Wikipedia, DBpedia, and YAGO. These open structural bases enrich large text and named entity that can be regarded as golden or silver standards of named entity. Richman and Schone [32] designed a system using Wikipedia to recognize and classify named entity and demonstrated that NER of other languages can benefit from the knowledge of English language. Pan et al. [4] developed a cross-lingual name labeling and linking structure for more than 200 languages with the help of a series of mining methods of Knowledge Base.

2) Transfer Learning: Transfer Learning that leverages source domains knowledge to construct target domain task can overcome the lack of labeled data and repeatedly design specific model. Although some systems do not explicitly point out the concept of Transfer Learning, we can still find significant Transfer Learning characteristics. Three papers published as follows: Guo et al. [33] presented a domain adaption method with latent semantic association to solve the data distribution difference between source domain and target domain. Zirikly and Hagiwara [34] proposed an approach to cross-lingual NER model transferring by introducing multilingual gazetteers generated with cross-lingual word representation mappings and graph propagation. This approach transfers English e-commerce domain model to Chinese and Spanish. Ni et al. [35] presented four complicated projection-based approaches between source language and target language. The first two approaches are used for encoding, the last two approaches used for decoding.

3) Joint Learning Based on Multitask: Joint task refers to the case that one system can simultaneously deal with NER and other task(s). For example, Chen et al. [36] proposed an integrated model that adopts mapping type ratio feature to jointly identify and align bilingual named entities between English and Chinese. Wang et al. [37] presented a graphical model that jointly performs bilingual NER tagging and word alignment by the combination of two unidirectional alignment models and two monolingual tagging models for Chinese and English.

IV. Conclusion

As one of the first and important stage in NLP pipeline, research on NER attracts moderate attention for the past two decades. Meanwhile, many concrete aspects of NER already have changed. In the following chapter, we present conclusions and give some perspectives.

First, multilingual NER systems get more attention. Considering the similarity between multiple target languages, multilingual NER systems reduce the cost that each target language builds one NER system separately. Second, even if English always has a dominant position, languages except English soar up. Low resource languages receive more and more attention. Whats more, building low resource languages NER system is inevitable in real-world scenarios. Third, with the development of computing power and the richness of corpus, models based on neural networks are widely been used for dealing rich resource NER tasks. One of the disadvantages is the models based on deep learning need large amount of corpus so that low resource languages and some domains that lack large amounts corpora could not get high performance. Knowledge Bases enhances current NER system performance, Transfer Learning can partly overcome the scarcity of labeled training data, so both of all become popular.

Limited by the availability of language resource, NER of target languages with different resource conditions adopt different approaches. For rich resource languages, NER systems extend into more spacious domains (medical notes, noisy text, e-commerce and so on). For low resource languages, reducing the needs of labeled data (active learning, distance learning) and transferring rich resource language knowledge to target task (transfer Learning, knowledge Base) are two major ideas. Enormous and high-quality labeled data has always been difficult for most languages, overcoming the scarcity of languages resource is the key. How to effectively use current resource and methods of well-studied languages to build new NER system may be very valuable for the next step of research.

References

[1] R. Grishman and B. Sundheim, “Message understanding conference-6: A brief history,” in *COLING 1996 Volume 1: The 16th International Conference on Computational Linguistics*, vol. 1, 1996.

[2] D. Nadeau and S. Sekine, “A survey of named entity recognition and classification,” *Lingvisticae Investigationes*, vol. 30, no. 1, pp. 3–26, 2007.

[3] D. D. Palmer and D. S. Day, “A statistical profile of the named entity task,” in *Proceedings of the fifth conference on Applied natural language processing*. Association for Computational Linguistics, 1997, pp. 190–193.

[4] X. Pan, B. Zhang, J. May, J. Nothman, K. Knight, and H. Ji, “Cross-lingual name tagging and linking for 282 languages,” in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, vol. 1, 2017, pp. 1946–1958.

[5] K. Riaz, “Rule-based named entity recognition in urdu,” in *Proceedings of the 2010 named entities workshop*. Association for Computational Linguistics, 2010, pp. 126–135.

[6] I. Demiros, S. Boutsis, V. Giouli, M. Liakata, H. Papageorgiou, and S. Piperidis, “Named entity recognition in greek texts.” in *LREC*, 2000.

[7] T. Iwakura, “A named entity recognition method using rules acquired from unlabeled data,” in *Proceedings of the International Conference Recent Advances in Natural Language Processing 2011*, 2011, pp. 170–177.

[8] A. De Amaral, “Rule-based named entity extraction for ontology population,” in *Proceedings of the Student Research Workshop associated with RANLP 2013*, 2013, pp. 58–62.

[9] R. Yeniterzi, “Exploiting morphology in turkish named entity recognition system,” in *Proceedings of the ACL 2011 Student Session*. Association for Computational Linguistics, 2011, pp. 105–110.

[10] U. Singh, V. Goyal, and G. S. Lehal, “Named entity recognition system for urdu,” *Proceedings of COLING 2012*, pp. 2507–2518, 2012.

[11] J. Piskorski and M. Ehrmann, “On named entity recognition in targeted twitter streams in polish.” in *Proceedings of the 4th Biennial International Workshop on Balto-Slavic Natural Language Processing*, 2013, pp. 84–93.

[12] A. McCallum and W. Li, “Early results for named entity recognition with conditional random fields, feature induction and web-enhanced lexicons,” in *Proceedings of the seventh conference on Natural language learning at HLT-NAACL 2003-Volume 4*. Association for Computational Linguistics, 2003, pp. 188–191.

[13] S. Sekine, “Nyu: Description of the japanese ne system used for met-2,” in *Proc. of the Seventh Message Understanding Conference (MUC-7*. Citeseer, 1998.

[14] D. M. Bikel, S. Miller, R. Schwartz, and R. Weischedel, “Nymble: a high-performance learning name-finder,” in *Proceedings of the fifth conference on Applied natural language processing*. Association for Computational Linguistics, 1997, pp. 194–201.

[15] A. Borthwick, J. Sterling, E. Agichtein, and R. Grishman, “Nyu: Description of the mene named entity system as used in muc-7,” in *Seventh Message Understanding Conference (MUC-7): Proceedings of a Conference Held in Fairfax, Virginia, April 29-May 1, 1998*, 1998.

[16] M. Asahara and Y. Matsumoto, “Japanese named entity extraction with redundant morphological analysis,” in *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1*. Association for Computational Linguistics, 2003, pp. 8–15.

[17] D. Wu, G. Ngai, M. Carpuat, J. Larsen, and Y. Yang, “Boosting for named entity recognition,” in *proceedings of the 6th conference on Natural language learning-Volume 20*. Association for Computational Linguistics, 2002, pp. 1–4.

[18] D. Wu, G. Ngai, and M. Carpuat, “A stacked, voted, stacked model for named entity recognition,” in *Proceedings of the seventh conference on Natural language learning at HLT-NAACL 2003-Volume 4*. Association for Computational Linguistics, 2003, pp. 200–203.

[19] W. Liao and S. Veeramachaneni, “A simple semi-supervised algorithm for named entity recognition,” in *Proceedings of the NAACL HLT 2009 Workshop on Semi-Supervised Learning for Natural Language Processing*. Association for Computational Linguistics, 2009, pp. 58–65.

[20] D. Küçük and R. Steinberger, “Experiments to improve named entity recognition on turkish tweets,” *arXiv preprint arXiv:1410.8668*, 2014.

[21] J. Brooke, A. Hammond, and T. Baldwin, “Bootstrapped text-level named entity recognition for literature,” in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, vol. 2, 2016, pp. 344–350.

[22] U. Kruschwitz and M. Poesio, “Combining minimally-supervised methods for arabic named entity recognition,” *Transactions of the Association of Computational Linguistics*, vol. 3, no. 1, pp. 243–255, 2015.

[23] M. Elsner, E. Charniak, and M. Johnson, “Structured generative models for unsupervised named-entity clustering,” in *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*. Association for Computational Linguistics, 2009, pp. 164–172.

[24] M. Bhagavatula, S. GSK, and V. Varma, “Language-independent named entity identification using wikipedia,” in *Proceedings of the First Workshop on Multilingual Modeling*. Association for Computational Linguistics, 2012, pp. 11–17.

[25] S. Patil, S. Pawar, and G. Palshikar, “Named entity extraction using information distance,” in *Proceedings of the Sixth International Joint Conference on Natural Language Processing*, 2013, pp. 1264–1270.

[26] J. P. Chiu and E. Nichols, “Named entity recognition with bidirectional lstm-cnns,” *arXiv preprint arXiv:1511.08308*, 2015.

[27] J. Hammerton, “Named entity recognition with long short-term memory,” in *Proceedings of the seventh conference on Natural language learning at HLT-NAACL 2003-Volume 4*. Association for Computational Linguistics, 2003, pp. 172–175.

[28] G. Lample, M. Ballesteros, S. Subramanian, K. Kawakami, and C. Dyer, “Neural architectures for named entity recognition,” *arXiv preprint arXiv:1603.01360*, 2016.

[29] Y. Chen, Y. Ouyang, W. Li, D. Zheng, and T. Zhao, “Using deep belief nets for chinese named entity categorization,” in *Proceedings of the 2010 Named Entities Workshop*. Association for Computational Linguistics, 2010, pp. 102–109.

[30] M. Xu, H. Jiang, and S. Watcharawittayakul, “A local detection approach for named entity recognition and mention detection,” in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, vol. 1, 2017, pp. 1237–1247.

[31] C. N. d. Santos and V. Guimaraes, “Boosting named entity recognition with neural character embeddings,” *arXiv preprint arXiv:1505.05008*, 2015.

[32] A. E. Richman and P. Schone, “Mining wiki resources for multilingual named entity recognition,” *Proceedings of ACL-08: HLT*, pp. 1–9, 2008.

[33] H. Guo, H. Zhu, Z. Guo, X. Zhang, X. Wu, and Z. Su, “Domain adaptation with latent semantic association for named entity recognition,” in *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*. Association for Computational Linguistics, 2009, pp. 281–289.

[34] A. Zirikly and M. Hagiwara, “Cross-lingual transfer of named entity recognizers without parallel corpora,” in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, vol. 2, 2015, pp. 390–396.

[35] J. Ni, G. Dinu, and R. Florian, “Weakly supervised cross-lingual named entity recognition via effective annotation and representation projection,” *arXiv preprint arXiv:1707.02483*, 2017.

[36] Y. Chen, C. Zong, and K.-Y. Su, “On jointly recognizing and aligning bilingual named entities,” in *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, 2010, pp. 631–639.

[37] M. Wang, W. Che, and C. D. Manning, “Joint word alignment and bilingual named entity recognition using dual decomposition,” in *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, vol. 1, 2013, pp. 1073–1082.