Experiments to Improve Named Entity Recognition on Turkish Tweets

Dilek Küçük and **Ralf Steinberger**
European Commission, Joint Research Centre
Via E. Fermi 2749
21027 Ispra (VA), Italy
`firstname.lastname@jrc.ec.europa.eu`

Abstract

Social media texts are significant information sources for several application areas including trend analysis, event monitoring, and opinion mining. Unfortunately, existing solutions for tasks such as named entity recognition that perform well on formal texts usually perform poorly when applied to social media texts. In this paper, we report on experiments that have the purpose of improving named entity recognition on Turkish tweets, using two different annotated data sets. In these experiments, starting with a baseline named entity recognition system, we adapt its recognition rules and resources to better fit Twitter language by relaxing its capitalization constraint and by diacritics-based expansion of its lexical resources, and we employ a simplistic normalization scheme on tweets to observe the effects of these on the overall named entity recognition performance on Turkish tweets. The evaluation results of the system with these different settings are provided with discussions of these results.

1 Introduction

Analysis of social media texts, particularly microblog texts like tweets, has attracted recent attention due to significance of the contained information for diverse application areas like trend analysis, event monitoring, and opinion mining. Tools for well-studied problems like named entity recognition (NER) are usually employed as components within these social media analysis applications. For instance, in (Abel et al., 2011), named entities extracted from tweets are used to determine trending topics for user modeling within the context of personalized recommender systems and

in (Ritter et al., 2012), named entities in tweets are used to complement the events extracted by an open domain event extraction system for Twitter. However, existing NER solutions for well-formed text types like news articles are reported to suffer from considerable performance degradations when they are ported to social media texts, mainly due to the peculiarities of this latter text type (Ritter et al., 2011).

In this paper, we report on our NER experiments on Turkish tweets in order to determine facilitating and impeding factors during the development of a NER system for Turkish tweets which can be used in social media analysis applications. We carry out these experiments on two tweet data sets annotated with named entities. After the initial evaluation results of a rule-based NER system (Küçük and Yazıcı, 2009) on these data sets, we gradually present the performance results achieved by the extended versions of the system together with discussions of these results. For these experiments, we first perform two system adaptations, i.e., relaxing the capitalization constraint of the system and diacritics-based expansion of the system's lexical resources. Next, we incorporate a simplistic tweet normalization scheme into the NER procedure. After the evaluation of these extensions, we provide discussions on the plausible features of a NER system tailored to Turkish tweets.

The rest of the paper is organized as follows: In Section 2, we review the literature on NER on tweets and NER on Turkish texts. In Section 3, we present our NER experiments on Turkish tweets. Directions of future work are outlined in Section 4 and finally Section 5 concludes the paper.

2 Related Work

There are several recent studies presenting approaches for NER on microblog texts, especially on tweets in English. Among these studies, in (Ritter et al., 2011), a NER system tailored to

tweets, called T-NER, is presented which employs Conditional Random Fields (CRF) for named entity segmentation and labelled topic modelling for subsequent classification, using Freebase dictionaries. A hybrid approach to NER on tweets is presented in (Liu et al., 2011) where k-Nearest Neighbor and CRF based classifiers are sequentially applied. In (Liu et al., 2012), a factor graph based approach is proposed that jointly performs NER and named entity normalization on tweets. An unsupervised approach that performs only named entity extraction on tweets using resources like Wikipedia is described in (Li et al., 2012). A clustering-based approach for NER on microtexts is presented in (Jung, 2012), a lightweight filter based approach for NER on tweets is described in (de Oliveira et al., 2013), and a series of NER experiments on targeted tweets in Polish is presented in (Piskorski and Ehrmann, 2013). Finally, an adaptation of the ANNIE component of GATE framework to microblog texts, called TwitIE, is described in (Bontcheva et al., 2013).

Considering NER research on Turkish texts, various approaches have been employed so far including those based on using Hidden Markov Models (HMM) (Tür et al., 2003), on manually engineered recognition rules (Küçük and Yazıcı, 2009; Küçük and Yazıcı, 2012), on rule learning (Tatar and Çicekli, 2011), and on CRFs (Yeniterzi, 2011; Şeker and Eryiğit, 2012). All of these approaches have been proposed for news texts and the CRF-based approach (Şeker and Eryiğit, 2012) is reported to outperform the previous proposals with a balanced F-Measure of about 91%.

To the best of our knowledge, there are only two studies on NER from Turkish tweets. In (Çelikkaya et al., 2013), the CRF-based NER system (Şeker and Eryiğit, 2012) is evaluated on informal text types and is reported to achieve an F-Measure of 19% on tweets. In (Küçük et al., 2014), a tweet data set in Turkish annotated with named entities is presented. The adaptation of a multilingual rule-based NER system (Pouliquen and Steinberger, 2009) to Turkish, which achieves an F-Measure of about 61% on a news article data set, gets an F-Measure of 37% on this tweet data set, and after extending the resources of the NER system with frequently appearing person and organization names in Turkish news articles, the corresponding scores increase to about 69% and 43%, respectively (Küçük et al., 2014).

Table 1: NE Statistics on the Data Sets.

NE Type	Frequency in	
	Tweet Set-1	Tweet Set-2
Person	457	774
Location	282	191
Organization	241	409
All PLOs	980	1,374
Date	201	342
Time	5	25
Money	16	13
Percent	9	3
All NEs	1,211	1,757

3 Named Entity Recognition Experiments

The NER experiments are performed using the rule-based NER system (Küçük and Yazıcı, 2009) which makes use of a set of lexical resources, i.e., lists of person/location/organization names (henceforth referred to as PLOs), and patterns for the extraction of named entities (NEs) of type PLOs and time/date/money/percent expressions (Küçük and Yazıcı, 2009). The system is proposed for news articles which is a considerably well-formed text type usually with proper capitalization of the initial letters of PLOs and separation of these PLOs from their suffixes with apostrophes[1]. Yet, as even such well-formed texts may be lacking these important indicators of PLOs, the system can be configured to make use of the capitalization clue or not, and it includes a simplistic morphological analyzer to check the suffixes at the end of PLO candidates and thereby validate these candidates (Küçük and Yazıcı, 2009).

This NER system achieves a balanced F-Measure of 78.7% (without giving any credit to partial extractions) on a news article data set of about 20K tokens obtained from the METU Turkish corpus (Say et al., 2002) where the annotated form of this data set includes a total of 1,613 NEs. Within the course of the current study, we have evaluated this system on two tweet data sets in Turkish where statistical information about these data sets are provided in Table 1. The first one, which is referred to as $Tweet\ Set-1$ in Table 1, is presented in (Küçük et al., 2014) and comprises 2,320 tweets with about 20K tokens. The second data set ($Tweet\ Set-2$) includes about 5K

---

[1]An example inflected named entity of location name type (a city name) in Turkish which takes the dative case suffix ($-ya$) is $Ankara'ya$ (meaning $to\ Ankara$) where the initial letter of the named entity is properly capitalized and the case suffix is accordingly separated from the entity with an apostrophe.

tweets with about 50K tokens and is described in (Çelikkaya et al., 2013).

3.1 Initial Experiments

We have first evaluated the system's performance on the data sets without any extensions to the existing NER system. Table 2 presents these evaluation results using the commonly employed metrics of precision, recall, and balanced F-Measure, without giving any credit to partially extracted NEs. Table 3 displays those results with the same metrics this time giving credit to partial extractions with the constraint that the NE type within the system output and the answer key must be the same, where these metrics have been employed in studies like (Maynard et al., 2001).

The evaluation results in Table 2 and Table 3 are in line with the common finding reported in the literature that the NER systems for comparatively well-formed text types face considerable performance decreases when they are evaluated on tweets. This observation is usually attributed to the peculiarities of tweet texts such as common grammatical/spelling errors and deliberate contractions. With strict metrics, the system is reported to achieve an F-Measure rate of 78.7%. When it is ported to tweets, the best overall F-Measure rates achieved are 53.23% and 44.25% on $Tweet\ Set-1$ and $Tweet\ Set-2$, respectively, while the corresponding best F-Measure rates for only PLOs are 47.76% and 36.63%, respectively, all with strict metrics. The difference between the results for PLOs and the overall results also confirms that the system recognizes temporal and numerical expressions (within its scope) with decent performance, compared to the recognition of PLOs.

The F-Measure rates obtained when partial extractions are also given credit are about 5% higher than those obtained without giving any credit to partially extracted NEs. This increase is important due to pragmatic reasons as these partially extracted NEs can help conveniently filter tweet streams and retrieve relevant subsets of tweets in several application settings.

3.2 NER Experiments with Rule/Resource Adaptations

Tweet texts possess the following peculiarities usually as opposed to other formal text types:

- Grammatical/spelling errors are common,

like incorrectly writing proper names all in lowercase letters. A Turkish example illustrating a spelling error is the use of $geliyoooo$ instead of $geliyor$ (meaning $is\ coming$).

- Contracted word forms are commonly used instead of full forms, like referring to the football club called $Fenerbahçe$ as $Fener$ only, where the latter contracted form is also homonymous to a common name in Turkish (meaning $lantern$).

- For the particular case of Turkish tweets, non-accentuated characters (c, g, i, o, s, and u) are often utilized instead of the corresponding Turkish characters with diacritics (ç, ğ, ı, ö, ş, and ü). An example of this phenomenon is writing $cunku$ instead of the correct form, $çünkü$ (meaning $because$).

Considering the above features, in order to improve the initial NER performance on Turkish tweets, we have tested two adaptations of the rule-based NER system. The details of these adaptations and the corresponding evaluation results are presented in the following subsections.

3.2.1 Relaxing the Capitalization Constraint of the System

As proper capitalization of PLOs is usually lacking in tweets, we have evaluated the NER system with its capitalization feature turned off, so that the system considers all tokens (no matter whether their initial character is capitalized or not) as valid NE candidates. The initial evaluation results of the system with this setting are provided in Table 2 and Table 3 within the rows where the $Capitalization$ column has a corresponding OFF value. The results for these two capitalization settings are also similarly provided in Tables 4-6 which present the evaluation results described in the upcoming sections.

The results in Table 2 and Table 3 demonstrate that relaxing the capitalization constraint (i.e., not using the capitalization clue) during the NER procedure on Turkish tweets consistently improves performance for PLOs on both data sets. The improvement obtained with this relaxation is more dramatic on $Tweet\ Set-2$ and for this data set the overall results are accordingly better than those obtained when the capitalization clue is used. It should again be noted that the NER system uses a

Table 2: Initial NER Evaluation Results (Strict Metrics).

Data Set	Capitalization	Metric	Person	Location	Organization	Overall for PLOs	Overall for 7 Types
Tweet Set-1	ON	P (%)	52.82	77.78	72.34	64.16	71.13
		R (%)	32.82	49.65	28.22	36.53	42.53
		F (%)	**40.49**	**60.61**	**40.60**	**46.55**	**53.23**
	OFF	P (%)	36.73	71.72	58.70	49.29	56.21
		R (%)	43.33	62.06	33.61	46.33	50.45
		F (%)	**39.76**	**66.54**	**42.74**	**47.76**	**53.18**
Tweet Set-2	ON	P (%)	55.79	58.68	72.06	58.86	65.62
		R (%)	20.54	37.17	11.98	20.31	30.85
		F (%)	**30.03**	**45.51**	**20.55**	**30.19**	**41.97**
	OFF	P (%)	35.61	45.53	40.72	38.31	46.27
		R (%)	38.37	61.26	16.63	35.08	42.40
		F (%)	**36.94**	**52.23**	**23.61**	**36.63**	**44.25**

Table 3: Initial NER Evaluation Results (Partial Metrics).

Data Set	Capitalization	Metric	Person	Location	Organization	Overall for PLOs	Overall for 7 Types
Tweet Set-1	ON	P (%)	65.33	86.05	88.37	75.98	80.74
		R (%)	39.38	54.01	32.34	41.87	47.13
		F (%)	**49.14**	**66.37**	**47.35**	**53.99**	**59.52**
	OFF	P (%)	42.83	78.68	69.11	56.25	62.49
		R (%)	50.92	67.71	38.00	52.55	55.72
		F (%)	**46.53**	**72.78**	**49.04**	**54.34**	**58.91**
Tweet Set-2	ON	P (%)	69.79	61.34	74.63	68.27	72.51
		R (%)	24.28	38.62	12.25	22.65	33.31
		F (%)	**36.03**	**47.40**	**21.05**	**34.02**	**45.65**
	OFF	P (%)	41.82	48.41	41.99	43.21	50.91
		R (%)	45.10	65.59	17.06	39.38	46.45
		F (%)	**43.40**	**55.71**	**24.26**	**41.21**	**48.58**

simplistic morphological analyzer to validate suffixes added at the ends of the NEs, thereby the system does not overgenerate with this new setting, although the precision rates decrease considerably in return to corresponding increases in the recall rates. To summarize, together with the fact that about 25.1% of all PLOs within $Tweet\ Set-1$ are lacking proper capitalization (Küçük et al., 2014), these findings suggest that the ability to relax this capitalization constraint is a convenient feature of a practical NER system for Turkish tweets. An alternative feature would be to automatically correct the capitalization of NEs instead, as a preprocessing step.

3.2.2 Diacritics-Based Expansion of the Lexical Resources

In Turkish tweet texts, words including Turkish characters with diacritics are often, usually either erroneously or deliberately for pragmatic reasons such as to type faster, spelled with their non-diacritic equivalents, as pointed out above. Therefore, we expand the entries in the lexical resources of the NER system to include both diacritic and non-diacritic variants of these entries. For instance, the Turkish name of the island $Cyprus$, $Kıbrıs$, may appear in tweets as $Kibrıs$, $Kıbris$, or $Kibris$, as well. As this example denotes, for each existing entry with n such Turkish-specific characters, 2^n entries (including the original entry) are included in the ultimate expanded forms

of the lexical resources, since each such character may be used as it is or may be replaced with its equivalent.

During this expansion stage, we have applied a filtering procedure over these newly considered $2^n - 1$ entries to check whether they are homonymous to common names in Turkish. This filtering procedure basically checks whether an expansion candidate is within a list of unique, supposedly well-formed, Turkish words comprising about 1,140,208 items including inflected forms (Zemberek, 2010), and if it is, then this candidate is discarded to avoid overgeneration during the actual NER procedure.

We have tested this new version of the system with expanded lexical resources and the corresponding evaluation results are provided in Table 4 and Table 5, using the strict and partial evaluation metrics, respectively. Both strict and partial evaluation results denote that the performance of the system is improved after this diacritics-based expansion of the system resources. The best results are obtained when this expansion is combined with the relaxation of the capitalization constraint, for PLOs on $Tweet\ Set-1$, and both for PLOs and all 7 NE types on $Tweet\ Set-2$. Similar to the points made in the previous section, this diacritics-based expansion scheme stands as a promising feature of an ultimate NER system for Turkish tweets, also considering the fact that

Table 4: NER Evaluation Results After Diacritics-Based Expansion of Resources (Strict Metrics).

Data Set	Capitalization	Metric	Person	Location	Organization	Overall for PLOs	Overall for 7 Types
Tweet Set-1	ON	P (%)	53.00	78.80	73.20	64.89	71.95
		R (%)	32.82	51.42	29.46	37.35	44.26
		F (%)	40.54	62.23	42.01	47.41	54.81
	OFF	P (%)	36.17	71.31	59.03	48.95	56.16
		R (%)	43.76	63.48	35.27	47.35	52.35
		F (%)	39.60	67.17	44.16	48.13	54.19
Tweet Set-2	ON	P (%)	58.22	58.73	70.67	60.20	67.29
		R (%)	22.87	38.74	12.96	22.13	34.89
		F (%)	32.84	46.69	21.90	32.36	45.95
	OFF	P (%)	36.80	44.61	32.43	37.61	46.24
		R (%)	43.41	62.83	17.60	38.43	47.64
		F (%)	39.83	52.17	22.82	38.01	46.93

Table 5: NER Evaluation Results After Diacritics-Based Expansion of Resources (Partial Metrics).

Data Set	Capitalization	Metric	Person	Location	Organization	Overall for PLOs	Overall for 7 Types
Tweet Set-1	ON	P (%)	65.58	87.46	88.76	76.81	81.44
		R (%)	39.38	56.12	33.62	42.80	48.98
		F (%)	49.21	68.37	48.77	54.97	61.17
	OFF	P (%)	42.21	79.17	69.00	56.02	62.49
		R (%)	51.56	69.85	39.70	53.88	57.90
		F (%)	46.42	74.22	50.40	54.93	60.11
Tweet Set-2	ON	P (%)	71.48	61.29	72.97	69.07	73.68
		R (%)	26.68	40.21	13.24	24.51	37.47
		F (%)	38.86	48.56	22.41	36.18	49.67
	OFF	P (%)	42.26	47.07	33.33	41.75	50.23
		R (%)	50.14	66.76	18.04	42.65	51.72
		F (%)	45.86	55.21	23.41	42.20	50.96

about 6.3% of all NEs in $Tweet\ Set-1$ are written in characters with missing diacritics. A plausible alternative to this feature would be to perform diacritics-based correction (or, normalization) as presented in studies like (Mihalcea, 2002) prior to the actual NER procedure. Similar approaches can be tested on tweets in other languages having common characters with diacritics.

3.3 Tweet Normalization

Tweet normalization has emerged as an important research problem (Han and Baldwin, 2011), the solutions to which can readily be used in systems for sentiment analysis and NER (as considered in studies such as (Liu et al., 2012)), among others. In order to observe the effects of normalization on NER performance on Turkish tweets, we have first experimented with a simplistic tweet normalization scheme which aims at decreasing repeated characters in words, as repetition of characters in tweets is a frequent means to express stress. The scheme is outlined below:

1. In order to determine the list of valid Turkish words with consecutively repeated characters, we have employed the list of Turkish unique words (Zemberek, 2010), that we have previously utilized during the diacritics-based resource expansion procedure in Section 3.2.2. Within this list, 74,262 words (about 6.5% of the list) turn out to have con-

secutively repeated characters.

2. Using this sublist as a reference resource, we have implemented the actual simplistic normalization scheme: if a word in a tweet has consecutively repeated character sequences and the word is not included within the aforementioned sublist, then all of these character sequences are contracted to single character instances. For instance, with this procedure, the token $zamaanlaaa$ is correctly replaced with $zamanla$ (meaning $with\ time$) and $mirayyy$ is correctly replaced with $miray$ (a proper person name).

The employment of the above normalization scheme prior to the actual NER procedure has led to slightly poorer results as some NEs which should not be normalized through this scheme are normalized instead. For instance, the city name $Çanakkale$ is changed to $Çanakale$ during the normalization procedure and it is missed by the subsequent NER procedure. Hence, we employ a three-phase pipelined NER approach where we first run the NER procedure on the input text, then employ the normalization scheme on the NER output, and finally run the NER procedure again on the normalization output, in order to avoid that the normalization step corrupts well-formed NEs that can readily be extracted by the system.

The performance of this ultimate NER pipeline, with the capitalization feature turned off during

both of the actual NER phases, is evaluated only on $Tweet\ Set-1$. Therefore, the performance evaluations of the first NER phase correspond to the previously presented results in the rows 4-6 of Table 2 and Table 3, with strict and partial versions of the metrics, respectively.

Below we summarize our findings regarding the intermediate normalization procedure employed, based on its evaluation results. Although some of these findings are not directly relevant for the purposes of the NER procedure, we provide them for the completeness of the discussion on the normalization of Turkish tweets.

- Excluding the normalization cases which involve non-alphabetical characters only (like normalizing >>>>>> to >), those that result in a normalized form with a single alphabetical character (like normalizing oooooo to o), and those that involve emotion expressions (like normalizing :DDDDD to :D), the number of resulting instances considered for performance evaluation is 494.

- The number of normalization instances in which an incorrect token is precisely converted into its corresponding valid form is 253, so, the precision of the overall normalization scheme is 51.21%.

- 117 of the incorrect cases are due to the fact that the token that is considered for normalization is a valid but foreign token (such as normalizing $Harry$ to $Hary$, $jennifer$ to $jenifer$, $full$ to ful, and $tweet$ to $twet$). Hence, these cases account for a decrease of 23.68% in the precision of the normalization scheme.

- 15 of the incorrect instances are due to the fact that Turkish characters with diacritics are not correctly used, hence they cannot be found within the reference sublist of valid Turkish words, and subsequently considered by the normalization procedure, although they could instead be subject to a diacritics-based normalization, as pointed out at the end of Section 3.2.2. For instance, şiir (meaning $poem$) is incorrectly written as $siir$ in a tweet and since it, in this incorrect form, cannot be found on the reference sublist, it is erroneously changed to sir. There are also

other incorrect instances in which superfluous characters are correctly removed with the normalization procedure, yet the resulting token is still not in its correct form as a subsequent diacritics-based correction is required. Though they are not considerably frequent (as we only consider here tokens with consecutively repeated characters), these instances serve to confirm that the restoration of diacritics should be considered along with other forms of normalization.

- Some other frequent errors made by the normalization scheme are due to incorrect tokenization as whitespaces to separate tokens can be missing due to writing errors or the tendency to write some phrases hashtag-like. An example case is incorrectly writing the adverb, $demek\ ki$ (meaning so or $that\ means$), as $demekki$ in a tweet, which in turn is erroneously changed to $demeki$ during normalization. This token, $demekki$, should not be considered within this type of normalization at all, although it needs processing to be transformed into its correct form, $demek\ ki$.

To summarize, the normalization scheme can be enhanced considering the above points, where proper treatment of non-Turkish tokens and the consideration of diacritics-based issues stand as the most promising directions of improvement. Other more elaborate ways of normalizing tweets, as presented in studies such as (Han and Baldwin, 2011), should also be tested together with the NER procedure, to observe their ultimate contribution. Along the way, a normalization dictionary for Turkish can be compiled, following studies like (Han et al., 2012).

The evaluation results of the ultimate three-phase NER pipeline are provided in Table 6, with the systems's capitalization feature turned off in both NER phases. Within the first three rows, the results with the strict evaluation metrics are displayed while the last three rows present those results obtained with the partial versions. When we examine the individual NER results after the incorporation of normalization scheme in details, we observe that there are cases where incorrectly normalizing some common names or slang/contracted words leads to them being extracted as NEs during the second NER phase. In order to prevent such

Table 6: Evaluation Results of the NER Pipeline with Normalization, on $Tweet\ Set-1$.

Metric Type	Metric	Person	Location	Organization	Overall for PLOs	Overall for 7 Types
Strict	P (%)	36.45	71.72	58.99	48.94	55.91
	R (%)	44.42	62.06	34.02	46.94	51.20
	F (%)	**40.04**	**66.54**	**43.16**	**47.92**	**53.45**
Partial	P (%)	42.32	78.68	69.35	55.73	62.04
	R (%)	52.07	67.71	38.43	53.18	56.48
	F (%)	**46.69**	**72.78**	**49.45**	**54.43**	**59.13**

false positives, the ways of improving the normalization procedure discussed above can be implemented and thereby less errors will be propagated into the second NER phase.

Though the overall results in Table 6 are slightly better than their counterparts when normalization is not employed, we cannot derive sound conclusions about the contribution of this normalization scheme to the overall NER procedure. The slight improvement is also an expected result as the size of the test data set is quite small and the number of NEs to be recognized after this type of normalization is already limited since only about 1% of all PLOs in $Tweet\ Set-1$ have incorrectly repeated consecutive characters. Yet, the results are still promising in that with a more elaborate normalization procedure evaluated on larger corpora, more dramatic increases in the NER performance can be obtained on Turkish tweets.

4 Future Work

Directions of future work based on the current study include the following:

- Following the points made throughout Section 3, several normalization schemes also involving case and diacritics restoration can be implemented and incorporated into the NER procedure on tweets.

- Since tweet texts are short and informal, they often lack contextual clues needed to perform an efficient NER procedure. Additionally, there is a tendency to mention new and popular NEs in tweets which might be missed by a NER system with static lexical resources. Hence, extending the lexical resources of the NER system with contemporary up-to-date NEs automatically obtained from Turkish news articles can be considered. For this purpose, we can readily employ resources like JRC-Names (Steinberger et al., 2011), a publicly available continuously-updated NE and name variant dictionary, as a source of up-to-date NEs in Turkish.

5 Conclusion

In this study, we target the problem of named entity recognition on Turkish tweets. We have carried out experiments starting with a rule-based recognition system and gradually extended it in two directions: adapting the rules/resources of the system and introducing a tweet normalization scheme into the recognition procedure. Thereby, we present our findings on named entity recognition on Turkish tweets in addition to those on the normalization of Turkish tweets. Based on these findings, we outline some desirable features of a named entity recognition system tailored to Turkish tweets. Future work includes the employment and testing of more elaborate tweet normalization procedures along the way, on larger tweet data sets, in addition to evaluating the system after its resources are automatically extended with dictionaries of up-to-date named entities.

Acknowledgments

References

Fabian Abel, Qi Gao, Geert-Jan Houben, and Ke Tao. 2011. Analyzing Temporal Dynamics in Twitter Profiles for Personalized Recommendations in the Social Web. In *Proceedings of the 3rd ACM International Web Science Conference*.

Kalina Bontcheva, Leon Derczynski, Adam Funk, Mark Greenwood, Diana Maynard, and Niraj Aswani. 2013. TwitIE: An Open-Source Information Extraction Pipeline for Microblog Text. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing*.

Gökhan Çelikkaya, Dilara Torunoğlu, and Gülşen Eryiğit. 2013. Named Entity Recognition on Real Data: A Preliminary Investigation for Turkish. In *Proceedings of the 7th International Conference on Application of Information and Communication Technologies*.

Gökhan A. Şeker and Gülşen Eryiğit. 2012. Initial Explorations on Using CRFs for Turkish Named

Entity Recognition. In *Proceedings of the International Conference on Computational Linguistics*, pages 2459–2474.

Diego Marinho de Oliveira, Alberto H.F. Laender, Adriano Veloso, and Altigran S. da Silva. 2013. FS-NER: A Lightweight Filter-Stream Approach to Named Entity Recognition on Twitter Data. In *Proceedings of the 22nd International Conference on World Wide Web Companion*, pages 597–604.

Bo Han and Timothy Baldwin. 2011. Lexical Normalisation of Short Text Messages: Makn Sens a #twitter. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*, pages 368–378.

Bo Han, Paul Cook, and Timothy Baldwin. 2012. Automatically Constructing a Normalisation Dictionary for Microblogs. In *Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*.

Jason J. Jung. 2012. Online Named Entity Recognition Method for Microtexts in Social Networking Services: A Case Study of Twitter. *Expert Systems with Applications*, 39(9):8066–8070.

Dilek Küçük and Adnan Yazıcı. 2009. Named Entity Recognition Experiments on Turkish Texts. In T. Andreasen et al., editor, *Proceedings of the International Conference on Flexible Query Answering Systems*, volume 5822 of *Lecture Notes in Computer Science*, pages 524–535.

Dilek Küçük and Adnan Yazıcı. 2012. A Hybrid Named Entity Recognizer for Turkish. *Expert Systems with Applications*, 39(3):2733–2742.

Dilek Küçük, Guillaume Jacquet, and Ralf Steinberger. 2014. Named Entity Recognition on Turkish Tweets. In *Proceedings of the Language Resources and Evaluation Conference*.

Chenliang Li, Jianshu Weng, Qi He, Yuxia Yao, Anwitaman Datta, Aixin Sun, and Bu-Sung Lee. 2012. TwiNER: Named Entity Recognition in Targeted Twitter Stream. In *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 721–730.

Xiaohua Liu, Shaodian Zhang, Furu Wei, and Ming Zhou. 2011. Recognizing Named Entities in Tweets. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1*, pages 359–367.

Xiaohua Liu, Ming Zhou, Furu Wei, Zhongyang Fu, and Xiangyang Zhou. 2012. Joint Inference of Named Entity Recognition and Normalization for Tweets. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers - Volume 1*, pages 526–535.

Diana Maynard, Valentin Tablan, Cristan Ursu, Hamish Cunningham, and Yorick Wilks. 2001. Named entity recognition from diverse text types. In *Proceedings of the Conference on Recent Advances in Natural Language Processing*.

Rada F. Mihalcea. 2002. Diacritics Restoration: Learning from Letters versus Learning from Words. In *Proceedings of the 3rd International Conference on Intelligent Text Processing and Computational Linguistics*, pages 339–348.

Jakub Piskorski and Maud Ehrmann. 2013. On Named Entity Recognition in Targeted Twitter Streams in Polish. In *Proceedings of the ACL Workshop on Balto-Slavic Natural Language Processing*.

Bruno Pouliquen and Ralf Steinberger. 2009. Automatic Construction of Multilingual Name Dictionaries. In C. Goutte et al., editor, *Learning Machine Translation*, Advances in Neural Information Processing Systems Series, pages 59–78. MIT Press.

Alan Ritter, Sam Clark, Mausam, and Oren Etzioni. 2011. Named Entity Recognition in Tweets: An Experimental Study. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 1524–1534.

Alan Ritter, Mausam, Oren Etzioni, and Sam Clark. 2012. Open Domain Event Extraction from Twitter. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1104–1112.

Bilge Say, Deniz Zeyrek, Kemal Oflazer, and Umut Özge. 2002. Development of a Corpus and a Treebank for Present-Day Written Turkish. In *Proceedings of the 11th International Conference of Turkish Linguistics*.

Ralf Steinberger, Bruno Pouliquen, Mijail Alexandrov Kabadjov, Jenya Belyaeva, and Erik Van der Goot. 2011. JRC-Names: A Freely Available, Highly Multilingual Named Entity Resource. In *Proceedings of the Conference on Recent Advances in Natural Language Processing*.

Serhan Tatar and İlyas Çicekli. 2011. Automatic Rule Learning Exploiting Morphological Features for Named Entity Recognition in Turkish. *Journal of Information Science*, 37(2):137–151.

Gökhan Tür, Dilek Hakkani-Tür, and Kemal Oflazer. 2003. A Statistical Information Extraction System for Turkish. *Natural Language Engineering*, 9(2):181–210.

Reyyan Yeniterzi. 2011. Exploiting Morphology in Turkish Named Entity Recognition System. In *Proceedings of the ACL Student Session*, pages 105–110.

Zemberek. 2010. Turkish Unique Word List of Zemberek NLP Library for Turkic Languages. Available at `http://zemberek.googlecode.com/files/full.txt.tr.tar.gz`.