Making Sense of Microposts (#Microposts2014) Named Entity Extraction & Linking Challenge

Amparo E. Cano
Knowledge Media Institue,
The Open University
amparo.cano@open.ac.uk

Giuseppe Rizzo
Università di Torino
EURECOM
giuseppe.rizzo@di.unito.it

Andrea Varga
The University of Sheffield
a.varga@dcs.shef.ac.uk

Matthew Rowe
Lancaster University
m.rowe@lancaster.ac.uk

Milan Stankovic
Sépage
Université Paris-Sorbonne
milstan@gmail.com

Aba-Sah Dadzie
The University of Birmingham
a.dadzie@cs.bham.ac.uk

ABSTRACT

Microposts are small fragments of social media content and a popular medium for sharing facts, opinions and emotions. They comprise a wealth of data which is increasing exponentially, and which therefore presents new challenges for the information extraction community, among others. This paper describes the 'Making Sense of Microposts' (#Microposts2014) Workshop's Named Entity Extraction and Linking (NEEL) Challenge, held as part of the 2014 World Wide Web conference (WWW'14). The task of this challenge consists of the automatic extraction and linkage of entities appearing within English Microposts on Twitter. Participants were set the task of engineering a named entity extraction and DBpedia linkage system targeting a predefined taxonomy, to be run on the challenge data set, comprising a manually annotated training and a test corpus of Microposts. 43 research groups expressed intent to participate in the challenge, of which 24 signed the agreement required to be given a copy of the training and test datasets. 8 groups fulfilled all submission requirements, out of which 4 were accepted for the presentation at the workshop and a further 2 as posters. The submissions covered sequential and joint methods for approaching the named entity extraction and entity linking tasks. We describe the evaluation process and discuss the performance of the different approaches to the #Microposts2014 NEEL Challenge.

Keywords

Microposts, Named Entity, Evaluation, Extraction, Linking, Disambiguation, Challenge

1. INTRODUCTION

Since the first Making Sense of Microposts (#MSM2011) workshop at the Extended Semantic Web Conference in 2011 through to the most recent workshop in 2014 we have received over 80 submissions covering a wide range of topics related to mining information and (re-)using the knowledge content of Microposts. Microposts are short text messages published using minimal effort via social media platforms. They provide a publicly available wealth of data which has proven to be useful in different applications and contexts (e.g. music recommendation, social bots, emergency response situations). However, gleaning useful information from Micropost content presents various challenges, due, among others, to the inherent characteristics of this type of data:

i) the limited length of Microposts;
ii) the noisy lexical nature of Microposts, where terminology differs between users when referring to the same thing, and abbreviations are commonplace.

A commonly used approach for mining Microposts is the use of cues that are available in textual documents, providing contextual features to this content. One example of such a cue is the use of named entities (NE). Extracting named entities in Micropost content has proved to be a challenging task; this was the focus of the first #MSM2013 challenge [3]. A step further into the use of such cues is to be able not only to recognize and classify them but also to provide further information, in other words, disambiguating entities. This prompted the Named Entity Extraction and Linking (NEEL) Challenge, held as part of the *Making Sense of Microposts Workshop (#Microposts2014)* at the *2014 World Wide Web Conference (WWW'14)*.

The purpose of this challenge was to set up an open and competitive environment that would encourage participants to deliver novel or improved approaches to extract entities from Microposts and link them to their DBpedia counterpart resources (if defined). This report describes the #Microposts2014 NEEL Challenge, our collaborative annotation of a corpus of Microposts and our evaluation of the performance of each submission. We also describe the approaches taken in the participants' systems – which use both established and novel, alternative approaches to entity extraction and linking. We describe how well they performed and how system performance differed across approaches. The resulting body of work has implications for researchers interested in the task of information extraction from social media.

2. THE CHALLENGE

In this section we describe the goal of the challenge, the task set, and the process we followed to generate the corpus of Microposts. We conclude the section with the list of the accepted submissions.

2.1 The Task and Goal

The NEEL Challenge task required participants to build semi-automated systems to:

i) extract entity mentions from a tweet. This stage is generally known as Named Entity Extraction (NEE);

ii) link each of these entities to an English DBpedia v3.9 resource. This stage is known as Named Entity Linking (NEL).

For this task we considered the definition of an *entity* in the general sense of being, in which an object or a set of objects do not necessarily need to have a material existence, but which however must be characterized as an instance of a taxonomy class. To facilitate the creation of the *gold standard* (GS) we limited the entity types evaluated in this challenge by specifying the taxonomy to be used: the NERD ontology v0.5[1] [16]. To this we added a few concepts from the DBpedia taxonomy. The taxonomy was not considered as normative in the evaluation of the submissions, nor for the ranking. This is a deliberate choice, to increase the complexity of the task and to let participants perform taxonomy matching starting from the distribution of the entities in the GS. The list of classes in the taxonomy used is distributed with the released GS[2].

Beside the typical word-tokens found in a Micropost, new to this year's challenge we considered special social media markers as entity mentions as well. These Twitter markers are tokens introduced with a special symbol. We considered two such markers: *hashtags*, prefixed by #, denoting the topic of a Micropost (e.g. *#londonriots*, *#surreyriots*, *#osloexpl*), and *mentions* prefixed by @, referring to Twitter user names, which include entities such as organizations (e.g. @*bbcworldservice*) and celebrities (e.g. @*ChadMMurray*, @*AmyWinehouse*).

Participants were required to recognize these different entity types within a given Micropost, and to extract the corresponding entity link tuples. Consider the following example, taken from our annotated corpus:

```
RT @bbcworldservice police confirms bomb in Oslo
#oslexp
```

the 2^{nd} token (the mention @*bbcworldservice*) in this Micropost refers to the international broadcaster, the BBC World Service; the 7^{th} token refers to the location *Oslo*; while the 8^{th} token (the hashtag #*oslexp*) refers to the 2011 Norway terrorist attack. An entry to the challenge would be required to spot these tokens and display them as an annotation with the following format:

```
bbcworldservice dbpedia:BBC_World_Service
Oslo dbpedia:Oslo
oslexp dbpedia:2011_Norway_attacks
```

where each line corresponds to a tab-separated *entity mention* and *entity link*[3]. We also consider the case where an entity is referenced in a tweet either as a noun or a noun phrase, if it:

a) belongs to one of the categories specified in the taxonomy;
b) is disambiguated by a DBpedia URI within the context of the tweet. Hence all entities without a disambiguation URI are not taken into account;
c) subsumes other entities

Therefore an entity phrase, composed of two or more entities, is considered a single entity if it can be disambiguated by a DBpedia URI. For our purposes, the longest entity phrase with a DBpedia URI will have precedence over shorter and single entities as in the following example:

1. [***Natural History Museum at Tring***];
2. [***News International chairman James Murdoch***]'s evidence to MPs on phone hacking;
3. [***Sony***]'s [***Android Honeycomb***] Tablet

For the 3^{nd} case, even though it may appear to be a coherent phrase, since there is no DBpedia URI for [*Sony's Android Honeycomb*], the entity phrase is split into valid component entities.

To encourage competition we solicited sponsorship for the winning submission. This was provided by the European project LinkedTV[4], who offered a prize of an iPad This generous sponsorship is testament to the growing interest in issues related to automatic approaches for gleaning information from (the very large amounts of) social media data.

2.2 Data Collection and Annotation

The challenge data set comprises 3,505 tweets extracted from a collection of over 18 million tweets. This collection, provided by the Redites project[5], covers event-annotated tweets collected for the period 15th July 2011 to 15th August 2011 (31 days). It extends over multiple notable events, including the death of Amy Winhehouse, the London Riots and the Oslo bombing. Since the NEEL Challenge task is to automatically extract and link entities, we built our data set considering both event and non-event tweets. Event tweets are more likely to contain entities; non-event tweets therefore enable us to evaluate the performance of the system in avoiding false positives in the entity extraction phase.

The data set was split into training (70%) and test (30%) sets. Statistics describing the training and test sets are provided in Table 1. The training set contains 2,340 tweets, totalling 41,037 tokens and 3,819 named entities; the test

[1] http://nerd.eurecom.fr/ontology/nerd-v0.5.n3
[2] The NEEL Challenge GS available for download from: http://ceur-ws.org/Vol-1141/microposts2014-neel_challenge_gs.zip
[3] In this example "dbpedia:" refers to the namespace prefix of a DBpedia resource (see http://dbpedia.org/resource)
[4] http://www.linkedtv.eu
[5] http://demeter.inf.ed.ac.uk/redites

Table 1: General statistics of the training and test data sets:
Posts **refers to the number of tweets in a data set;** ***Words*** **to the unique number of words;** ***Tokens*** **refers to the total number of words;** ***AvgTokens/Post*** **represents the average number of tokens per tweet;** ***NEs*** **denotes the unique number of NEs;** ***totalNEs*** **the total number of NEs; and** ***AvgNE/Post*** **the average number of NEs per post. We computed AvgTokens/Post and AvgNE/Post as the standard standard deviation from the mean (mean ± standard deviation).**

Data set	Posts	Words/Tokens	AvgTokens/Post	NEs	totalNEs	AvgNE/Post
train	2,340	12,758/41,037	17.54±5.70	1,862	3,819	3.26±3.37
test	1,165	6,858/20,224	17.36 ± 5.59	834	1,458	2.50±2.94

set contains 1,165 tweets, totalling 20,224 tokens and 1,458 named entities. The tweets are relatively long in both data sets, the average number of tokens per tweet is 17.54±5.70 in the train, and 17.36±5.59 in the test set. The average number of entities per tweet is also relatively high 3.26±3.37 for the training, and 2.50±2.94 for the test data set. The percentage of tweets without any entity is 32% (775 tweets) in the training, and 40% (469 tweets) in the test set. There is a fair bit of overlap of entities between the training and test data: 13.27% (316) of the named entities from the training data also occurs in the test data. Concerning the entities derived from the hashtag and mention social media markers, a total of 406 hashags were marked as entities in the training, and 184 were marked in the test set. The total frequency of entity mentions was 133 in the training, and 73 in the test data set.

The annotation of each Micropost in the training set gave all participants a common base from which to learn extraction patterns. In order to assess the performance of the submissions we used an underlying *gold standard* (GS), generated by 14 annotators, who had different backgrounds, including computer scientists, social scientists, social web experts, semantic web experts and linguists.

The annotation process was composed of the following phases:

Phase 1. Unsupervised annotation of the corpus was performed, to extract candidate links that were used as input to the next stage. The candidates were extracted using the NERD framework [15].

Phase 2. The data set was divided into batches so as to assign three different annotators to each batch. In this phase annotations were performed using CrowdFlower[6]. The annotators were asked to analyze the NERD links generated in phase 1 by adding or removing entity-annotations as they considered suitable. The annotators were also asked to mark any ambiguous cases encountered.

Phase 3. In the final, consistency-check, stage, three experts double-checked the collected annotations and generated the GS (for both the training and test sets). Three main tasks were carried out: (1) cross-consistency check of entity types; (2) cross-consistency check of URIs; (3) resolution of ambiguous cases raised by the 14 annotators. [7]

[6]http://crowdflower.com

The complete data set, including a list of changes and the gold standard, is available for download[8] with the #Microposts2014 Workshop proceedings, accessible under the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License[9].

2.3 Challenge Submissions

The challenge attracted a lot of interest from research groups spread across the world. Initially, 43 groups expressed their intent to participate in the challenge; however only 8 completed submission. Each submission consisted of a short paper explaining the system approach, and up to three different test set annotations generated by running the system with different settings. After peer review, 4 submissions were accepted, and a further 2 as posters. The submission run with the best overall performance for each system was used in the rankings (see Table 4). The submissions accepted are listed in Table 2.

2.4 System Descriptions

We present next an analysis of the participants' systems for the Named Entity Extraction and Linking (NEEL) tasks. Except for submission 18, who treated the NEEL task as a joint task of Named Entity Extraction (NEE) and Named Entity Linking (NEL); all participants approached the NEEL task as two sequential sub-tasks (i.e. NEE first, followed by NEL). A summary of these approaches includes:

i) use of *external systems*;

ii) *main features used*;

iii) type of *strategy* used;

iv) use of *external sources*.

Table 3 provides a detailed summary of the approaches used for both the NEE and NEL tasks.

[7]We aim to provide a more detailed explanation of the NEEL Challenge evaluation process in a separate publication.
[8]http://ceur-ws.org/Vol-1141/microposts2014-neel_challenge_gs.zip
[9]Following the Twitter ToS we only provide tweet IDs and annotations for the training set; and tweet IDs for the test set.

Table 2: Submissions accepted, ordered by submission number, with team affiliations and number of runs for each.

ID	Team's Affiliation	Authors	No. of runs
13	Twente	Habib M. et al.	2
15	Max Planck	Amir M. et al.	3
16	Hyberabad	Bansal R. et al.	1
18	Microsoft	Chang M.	3
19	Net7-Spaziodati-Pisa	Scaiella U. et al.	2
20	SAP	Dahlmeier D. et al.	1

The NEE task on Microposts is on its own challenging. One of the main *strategies* was to use off-the-shelf named entity recognition (NER) tools, improved through the use of extended gazetteers. System 18 approached the NEE task from scratch using a rule-based approach; all others made use of *external toolkits*. Some of these were Twitter-tuned and were applied for:

i) feature extraction, including the use of the TwitterNLP (2013) [13] and TwitterNLP (2011) [8] toolkits for POS tagging (systems 16, 20);

ii) entity extraction with TwiNER [10], Ritter's NER [14] and TAGME [5] (systems 13, 16, 19).

Other *external toolkits* which address NEE in longer newswire texts were also applied, including Stanford NER [6] and DBpedia Spotlight [11] (systems 15, 20).

Another common trend across these systems was the use of gazetteer-based, rule-matching approaches to improve the coverage of the off-the-shelf tools. System 13 applied simple regular expression rules to detect additional named entities not found by the NE extractor (such as numbers, and dates); systems 15 and 18 applied rules to find candidate entity mentions using a knowledge base (among others, Freebase [2]). Some systems also applied name normalization for feature extraction (systems 15, 18). This strategy was particularly useful for catering for entities originally appearing as hashtags or username mentions. For example, hashtags such as #BarackObama were normalized into a composite entity mention "Barack Obama"; and "@EmWatson" into "Emma Watson".

The NEL task involved in some cases the use of off-theself tools, for finding candidate links for each entity mention and/or for deriving mention features (systems 13, 19, 20). A common trend across systems was the use of external knowledge sources including:

i) NER dictionaries (e.g. Google CrossWiki [17]);

ii) Knowledge Base Gazetteers (e.g. Yago [9], DBpedia [1]);

iii) Weighted lexicons (using e.g. Freebase [2], Wikipedia);

iv) other sources (e.g. Microsoft Web N-gram [19]).

A wide range of different *features* were investigated for the linking strategies. Some systems characterized an entity using Micropost-derived features with Knowledge base (KB)-derived features (systems 13, 15, 16, 19). Micropost-derived features included the use of lexical (e.g., N-grams, capitalization) and syntactical (e.g., POS) features, while KB-derived features included the use of URIs, anchor text and link-based probabilities (see Table 3). Additionally, features were extended by capturing jointly the local (within a Micropost) and global (within a knowledge base) contextual information of an entity, via graph-based features (such as entity semantic cohesiveness) (system 18). Further novel features included the use of Twitter account metadata for characterizing mentions and popularity-based statistical features for characterizing entities (systems 16, 18).

The classification *strategies* used for entity linking included supervised approaches (systems 13, 15, 16, 18, 19) existing off-the-shelf approaches enhanced with simple heuristics (e.g. the search+rules) (system 20).

3. EVALUATION OF CHALLENGE SUBMISSIONS

We describe next the evaluation measures used to assess the goodness of the submissions and conclude with the final challenge rankings, with submissions ordered according to the F_1 measure.

3.1 Evaluation Measures

We evaluate the goodness of a system S in terms of the performance of the system to both recognize and link an entity from a test set TS. Per each instance in TS, a system provides a set of pairs P of the form: entity mention (e), and link (l). A link is any valid DBpedia URI[10] that points to an existing resource (e.g. `http://dbpedia.org/resource/Barack_Obama`). The evaluation consists of comparing submission entry pairs against those in the gold standard GS. The measures used to evaluate each pair are precision P, recall R, and f-measure F_1. The evaluation is based on micro-averages.

First, a cleansing stage is performed over each submission, resolving where needed, the redirects. Then, to assess the correctness of the pairs provided by a system S, we perform an exact-match evaluation, in which a pair is correct only if both the entity mention and the link match the corresponding set in the GS. Pair order is also relevant. We define $(e,l)_S \in S$ as the set of pairs extracted by the system S, $(e,l)_{GS} \in GS$ denotes the set of pairs in the gold standard. We define the set of true positives TP, false positives FP, and false negatives FN for a given system as:

$$TP = \{(e,l)_S | (e,l)_{GS} \in (S \cap GS)\} \tag{1}$$

$$FP = \{(e,l)_S | (e,l)_{GS} \in S \wedge (e,l) \notin GS)\} \tag{2}$$

$$FN = \{(e,l)_S | (e,l)_{GS} \in GS \wedge (e,l) \notin S\} \tag{3}$$

Thus TP defines the set of relevant pairs in TS, in other words the set of pairs in TS that match corresponding ones

[10] We consider all DBpedia v3.9 resources valid.

Table 3: Automated approaches for #Microposts2014 NEEL Challenge

Named Entity Extraction (NEE)

	External System	Main Features	NE Extraction Strategy	Linguistic Knowledge
13 UTwente	TwiNER [10]	Regular Expression, Entity phrases, N-gram	TwiNER [10] + CRF	DB[a] Gazetteer [1], Wiki[b] Gazetteer
15 MaxPlanck	StanfordNER [6]	-	-	NER Dictionary
16 IIT Hyberabad	RitterNER [14],TwitterNLP(2011) [8]	Proper nouns sequence, N-grams	-	Wiki
18 Microsoft	-	N-grams, stop words removal, punctuation as tokens	Rule-based (candidate filter)	Wiki and Freebase [2] lexicons
19 Net7-Spaziodati-UPisa	TAGME [5]	Wiki anchor texts, N-grams	Collective agreement + Wiki stats	Wiki
20 SAP	DBSpotlight [11], TwitterNLP(2013) [13]	Unigram, POS, lower, title & upper case, stripped words, isNumber, word cluster, DBpedia	CRF	DB Gazetteer, Brown Clusters [18]

Named Entity Linking (NEL)

	External System	Main Features	Linking Strategy	Linguistic Knowledge
13 UTwente	Google Search[c]	N-grams, DB links, Wiki links, Capitalization	SVM	Wiki, DB, WordNet [12], Web N-Gram [19], Yago [9]
15 MaxPlanck	-	Prefix, POS, suffix, Twitter account metadata, normalized mentions, tri-grams	entity aggregate prior + prefix-tree data structure + DB match	Wiki, DB, Yago [9]
16 IIT Hyberabad	-	Wiki context-based measure, anchor text measure, entity popularity (in Twitter)	LambdaMART [20] (ranking/disambiguation)	Wiki Gazetteer, Google CrossWiki Dictionary [17]
18 Microsoft	-	N-grams, lower case, entity graph features (entity semantic cohesiveness), popularity-based statistical features (clicks and visiting information from the Web)	DCD-SSVM [4] + MART gradient boosting [7]	Wiki, Freebase [2]
19 Net7-Spaziodati-UPisa	TAGME [5]	Link probability, mention-link commonness	C4.5 (for taxonomy-filter)	Wiki, DB [1]
20 SAP	Search API (Wiki, DBSpotlight [11], Google)	Entity mention	Search+rules	Wiki, DB [1]

[a] DBpedia abbreviated to ‘DB’ [1]
[b] Wikipedia abbreviated to ‘Wiki’
[c] Google Search, https://developers.google.com/web-search/docs

in GS. FP is the set of irrelevant pairs in TS, in other words the pairs in TS that do not match the pairs in GS. FN is the set of false negatives denoting the pairs that are not recognised by TS, yet appear in GS. Since our evaluation is based on a micro-average analysis, we sum the individual true positives, false positives, and false negatives of each system across all Microposts. As we require an exact-match for pairs (e, l) we are looking for strict entity recognition and linking matches; each system has to link each entity e recognised to the correct resource l.

From this set of definitions, we define precision, recall, and f-measure as follows:

$$P = \frac{|TP|}{|TP \cup FP|} \tag{4}$$

$$R = \frac{|TP|}{|TP \cup FN|} \tag{5}$$

$$F_1 = 2 * \frac{P * R}{P + R} \tag{6}$$

The evaluation framework used in the challenge is available at `https://github.com/giusepperizzo/neeleval`.

3.2 Evaluation Results

Table 4 reports the performance of participants' systems, using the best run for each. The ranking is based on the F_1.

System 18 clearly outperformed other systems, with F_1 more than 15% higher than the next best system. System 18 differed from all other systems, by using a joint approach to the NEEL task. The others each divided the task into a sequential entity extraction and linking task. The approach in System 18 made use of features which capture jointly an entity's local and global contextual information, resulting in the best approach submitted to the #Microposts2014 NEEL Challenge.

4. CONCLUSIONS

The aim of the #Microposts2014 Named Entity Extraction & Linking Challenge was to foster an open initiative that would encourage participants to develop novel approaches for extracting and linking entity mentions appearing in Microposts. The NEEL task involved the extraction of entity mentions in Microposts and the linking of these entity mentions to DBpedia resources (where such exist).

Our motivation for hosting this challenge is the increased availability of third-party entity extraction and entity linking tools. Although such tools have proven to be a good starting point for entity linking, even for Microposts, the evaluation results show that the NEEL task remains challenging when applied to social media content with its peculiarities when compared to standard length text.

As a result of this challenge, and the collaboration of annotators and participants, we also generated a manually annotated data set, which may be used in conjunction with the NEEL evaluation framework (`neeleval`). To the best of our knowledge this is the largest publicly available data set providing entity/resource annotations for Microposts. We hope that both the data set and the `neeleval` framework will facilitate the development of future approaches in this and other such tasks.

The results of this challenge highlighted the relevance of normalization and time-dependent features (such as popularity) for dealing with this type of progressively changing content. It also indicated that learning entity extraction and linking as a joint task may be beneficial for boosting performance in entity linking in Microposts.

We aim to continue to host additional challenges targeting more complex tasks, within the context of data mining of Microposts.

5. ACKNOWLEDGMENTS

The authors thank Nikolaos Aletras for helping to set-up the crowdflower experiments for annotating the gold standard data set. A special thank to Andrés García-Silva, Daniel Preoţiuc-Pietro, Ebrahim Bagheri, José M. Morales del Castillo, Irina Temnikova, Georgios Paltoglou, Pierpaolo Basile and Leon Derczynski, who took part in the annotation tasks. We also thank the participants who helped us improve the gold standard. We finally thank the LinkedTV project for supporting the challenge by sponsoring the prize for the winning submission.

6. REFERENCES

[1] S. Auer, C. Bizer, G. Kobilarov, J. Lehmann, and Z. Ives. DBpedia: A Nucleus for a Web of Open Data. In *6th International Semantic Web Conference (ISWC'07)*, 2007.

[2] K. Bollacker, C. Evans, P. Paritosh, T. Sturge, and J. Taylor. Freebase: A collaboratively created graph database for structuring human knowledge. In *ACM SIGMOD International Conference on Management of Data (SIGMOD'08)*, 2008.

[3] A. E. Cano Basave, A. Varga, M. Rowe, M. Stankovic, and A.-S. Dadzie. Making Sense of Microposts (#MSM2013) Concept Extraction Challenge. In *Making Sense of Microposts (#MSM2013) Concept Extraction Challenge*, 2013.

[4] M.-W. Chang and W.-T. Yih. Dual coordinate descent algorithms for efficient large margin structured prediction. *Transactions of the Association for Computational Linguistics*, 2013.

[5] P. Ferragina and U. Scaiella. Fast and accurate annotation of short texts with Wikipedia pages. *IEEE Software*, 29(1):70–75, 2012.

[6] J. R. Finkel, T. Grenager, and C. Manning. Incorporating non-local information into information extraction systems by Gibbs sampling. In *43rd Annual Meeting on Association for Computational Linguistics (ACL'05)*, 2005.

[7] J. H. Friedman. Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29:1189–1232, 2000.

[8] K. Gimpel, N. Schneider, B. O'Connor, D. Das, D. Mills, J. Eisenstein, M. Heilman, D. Yogatama, J. Flanigan, and N. A. Smith. Part-of-speech tagging for Twitter: Annotation, features, and experiments. In *49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, 2011.

Table 4: **P, R, F_1 breakdown figures per submission.**

Rank	System	Entry	P	R	F_1
1	18-2	Microsoft	77.10	64.20	70.06
2	13-2	Twente	57.30	52.74	54.93
3	19-2	Net7-Spaziodati-Pisa	60.93	42.25	49.90
4	15-3	MaxPlanck	53.28	39.51	45.37
5	16-1	Hyberabad	50.95	40.67	45.23
6	20-1	SAP	49.58	32.17	39.02

[9] J. Hoffart, F. M. Suchanek, K. Berberich, E. Lewis-Kelham, G. de Melo, and G. Weikum. YAGO2: Exploring and querying world knowledge in time, space, context, and many languages. In *20^{th} International Conference Companion on World Wide Web (WWW'11)*, 2011.

[10] C. Li, J. Weng, Q. He, Y. Yao, A. Datta, A. Sun, and B.-S. Lee. TwiNER: Named entity recognition in targeted Twitter stream. In *35^{th} International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '12)*, 2012.

[11] P. N. Mendes, M. Jakob, A. García-Silva, and C. Bizer. DBpedia spotlight: shedding light on the web of documents. In *7^{th} International Conference on Semantic Systems (I-Semantics'11)*, 2011.

[12] G. A. Miller. Wordnet: A lexical database for English. *Commun. ACM*, 38(11):39–41, 1995.

[13] O. Owoputi, C. Dyer, K. Gimpel, N. Schneider, and N. A. Smith. Improved part-of-speech tagging for online conversational text with word clusters. In *NAACL*, 2013.

[14] A. Ritter, S. Clark, Mausam, and O. Etzioni. Named entity recognition in tweets: An experimental study. In *EMNLP*, 2011.

[15] G. Rizzo and R. Troncy. NERD: A framework for unifying named entity recognition and disambiguation extraction tools. In *13^{th} Conference of the European Chapter of the Association for computational Linguistics (EACL'12)*, 2012.

[16] G. Rizzo, R. Troncy, S. Hellmann, and M. Bruemmer. NERD meets NIF: Lifting NLP extraction results to the Linked Data Cloud. In *Proceedings of the 5^{th} International Workshop on Linked Data on the Web (LDOW'12)*, 2012.

[17] V. I. Spitkovsky and A. X. Chang. A cross-lingual dictionary for English Wikipedia concepts. In *LREC*, 2012.

[18] J. Turian, L. Ratinov, and Y. Bengio. Word representations: A simple and general method for semi-supervised learning. In *48^{th} Annual Meeting of the Association for Computational Linguistics (ACL'10)*, 2010.

[19] K. Wang, C. Thrasher, E. Viegas, X. Li, and B.-j. P. Hsu. An overview of Microsoft Web N-gram corpus and applications. In *NAACL HLT 2010 Demonstration Session (HLT-DEMO'10)*, 2010.

[20] Q. Wu, C. J. Burges, K. M. Svore, and J. Gao. Adapting boosting for information retrieval measures. *Information Retrieval*, 13(3):254–270, 2010.