

Overview of the EVALITA 2016 Named Entity rEcognition and Linking in Italian Tweets (NEEL-IT) Task

Pierpaolo Basile¹ and Annalina Caputo² and Anna Lisa Gentile³ and Giuseppe Rizzo⁴

¹ Department of Computer Science, University of Bari Aldo Moro, Bari (Italy)

² ADAPT Centre, Trinity Collge Dublin, Dublin (Ireland)

³ University of Mannheim, Mannheim (Germany)

⁴ Istituto Superiore Mario Boella, Turin (Italy)

¹ pierpaolo.basile@uniba.it

² annalina.caputo@adaptcentre.ie

³ annalisa@informatik.uni-mannheim.de

⁴ giuseppe.rizzo@ismb.it

Abstract

English. This report describes the main outcomes of the 2016 Named Entity rEcognition and Linking in Italian Tweet (NEEL-IT) Challenge. The goal of the challenge is to provide a benchmark corpus for the evaluation of entity recognition and linking algorithms specifically designed for noisy and short texts, like tweets, written in Italian. The task requires the correct identification of entity mentions in a text and their linking to the proper named entities in a knowledge base. To this aim, we choose to use the canonicalized dataset of DBpedia 2015-10. The task has attracted five participants, for a total of 15 runs submitted.

Italiano. *In questo report descriviamo i principali risultati conseguiti nel primo task per la lingua Italiana di Named Entity rEcognition e Linking in Tweet (NEEL-IT). Il task si prefigge l'obiettivo di offrire un framework di valutazione per gli algoritmi di riconoscimento e linking di entità a nome proprio specificamente disegnati per la lingua italiana per testi corti e rumorosi, quali i tweet. Il task si compone di una fase di riconoscimento delle menzioni di entità con nome proprio nel testo e del loro successivo collegamento alle opportune entità in una base di conoscenza. In questo task abbiamo scelto come base di conoscenza la versione canonica di DBpedia 2015. Il task ha attirato cinque partecipanti per un totale di 15 diversi run.*

1 Introduction

Tweets represent a great wealth of information useful to understand recent trends and user behaviours in real-time. Usually, natural language processing techniques would be applied to such pieces of information in order to make them machine-understandable. Named Entity rEcognition and Linking (NEEL) is a particularly useful technique aiming to automatically annotate tweets with named entities. However, due to the noisy nature and shortness of tweets, this technique is more challenging in this context than elsewhere. International initiatives provide evaluation frameworks for this task, e.g. the Making Sense of Microposts workshop (Dadzie et al., 2016) hosted the 2016 NEEL Challenge (Rizzo et al., 2016), or the W-NUT workshop at ACL 2015 (Baldwin et al., 2015), but the focus is always and strictly on the English language. We see an opportunity to (i) encourage the development of language independent tools for Named Entity Recognition (NER) and Linking (NEL) systems and (ii) establish an evaluation framework for the Italian community. NEEL-IT at EVALITA has the vision to establish itself as a reference evaluation framework in the context of Italian tweets.

2 Task Description

NEEL-IT followed a setting similar to NEEL challenge for English Micropost on Twitter (Rizzo et al., 2016). The task consists of annotating each named entity mention (like *people*, *locations*, *organizations*, and *products*) in a text by linking it to a knowledge base (DBpedia 2015-10).

Specifically, each task participant is required to:

1. Recognize and typing each entity mention that appears in the text of a tweet;

Table 1: Example of annotations.

id	begin	end	link	type
288...	0	18	NIL	Product
288...	73	86	http://dbpedia.org/resource/Samsung_Galaxy_Note_II	Product
288...	89	96	http://dbpedia.org/resource/Nexus_4	Product
290...	1	15	http://dbpedia.org/resource/Carlotta_Ferlito	Person

- Disambiguate and link each mention to the canonicalized DBpedia 2015-10, which is used as referent Knowledge Base. This means that if an entity is present in the Italian DBpedia but not in the canonicalized version, this mention should be tagged as NIL. For example, the mention *Agorà* can only be referenced to the Italian DBpedia entry *Agorà <programma televisivo>*¹, but this entry has no correspondence into the canonicalized version of DBpedia. Then, it has been tagged as a NIL instance.
- Cluster together the non linkable entities, which are tagged as NIL, in order to provide a unique identifier for all the mentions that refer to the same named entity.

In the annotation process, a named entity is a string in the tweet representing a proper noun that: 1) belongs to one of the categories specified in a taxonomy and/or 2) can be linked to a DBpedia concept. This means that some concepts have a NIL DBpedia reference².

The taxonomy is defined by the following categories:

Thing languages, ethnic groups, nationalities, religions, diseases, sports, astronomical objects;

Event holidays, sport events, political events, social events;

Character fictional character, comics character, title character;

Location public places, regions, commercial places, buildings;

Organization companies, subdivisions of companies, brands, political parties, government

¹[http://it.dbpedia.org/resource/Agor%C3%A0_\(programma_televisivo\)](http://it.dbpedia.org/resource/Agor%C3%A0_(programma_televisivo))

²These concepts belong to one of the categories but they have no corresponding concept in DBpedia

bodies, press names, public organizations, collection of people;

Person people's names;

Product movies, tv series, music albums, press products, devices.

From the annotation are excluded the preceding article (like *il*, *lo*, *la*, etc.) and any other prefix (e.g. *Dott.*, *Prof.*) or post-posed modifier. Each participant is asked to produce an annotation file with multiple lines, one for each annotation. A line is a tab separated sequence of tweet id, start offset, end offset, linked concept in DBpedia, and category. For example, given the tweet with id 288976367238934528:

Chameleon Launcher in arrivo anche per smartphone: video beta privata su Galaxy Note 2 e Nexus 4: Chameleon Laun...

the annotation process is expected to produce the output as reported in Table 1.

The annotation process is also expected to link Twitter mentions (@) and hashtags (#) that refer to a named entities, like in the tweet with id 290460612549545984:

@CarlottaFerlito io non ho la forza di alzarmi e prendere il libro! Help me

the correct annotation is also reported in Table 1.

Participants were allowed to submit up to three runs of their system as TSV files. We encourage participants to make available their system to the community to facilitate reuse.

3 Corpus Description and Annotation Process

The NEEL-IT corpus consists of both a development set (released to participants as training set) and a test set. Both sets are composed by two TSV files: (1) the tweet id file, this is a list of all tweet ids used for training; (2) the gold standard,

containing the annotations for all the tweets in the development set following the format showed in Table 1.

The development set was built upon the dataset produced by Basile et al. (2015). This dataset is composed by a sample of 1,000 tweets randomly selected from the TWITA dataset (Basile and Nissim, 2013). We updated the gold standard links to the canonicalized DBpedia 2015-10. Furthermore, the dataset underwent another round of annotation performed by a second annotator in order to maximize the consistency of the links. Tweets that presented some conflicts were then resolved by a third annotator.

Data for the test set was generated by randomly selecting 1,500 tweets from the SENTIPOLC test data (Barbieri et al., 2016). From this pool, 301 tweets were randomly chosen for the annotation process and represents our Gold Standard (GS). This sub-sample was choose in coordination with the task organisers of SENTIPOLC (Barbieri et al., 2016), POSTWITA (Tamburini et al., 2016) and FacTA (Minard et al., 2016b) with the aim of providing a unified framework for multiple layers of annotations.

The tweets were split in two batches, each of them was manually annotated by two different annotators. Then, a third annotator intervened in order to resolve those debatable tweets with no exact match between annotations. The whole process has been carried out by exploiting BRAT³ web-based tool (Stenetorp et al., 2012).

Table 2 reports some statistics on the two sets: in both the most represented categories are “Person”, “Organization” and “Location”. “Person” is also the most populated category among the NIL instances, along to “Organization” and “Product”. In the development set, the least represented category is “Character” among the NIL instances and both “Thing” and “Event” between the linked ones. A different behaviour can be found in the test set where the least represented category is “Thing” in both NIL and linked instances.

4 Evaluation Metrics

Each participant was asked to submit up to three different run. The evaluation is based on the following three metrics:

STMM (*Strong Typed Mention Match*). This metrics evaluates the micro average F-1 score

³<http://brat.nlplab.org/>

Table 2: Datasets Statistics.

Stat.	Dev. Set	Test Set
# tweets	1,000	301
# tokens	14,242	4,104
# hashtags	250	108
# mentions	624	181
Mean token per tweet	14.24	13.65
# NIL Thing	14	3
# NIL Event	9	7
# NIL Character	4	5
# NIL Location	6	9
# NIL Organization	49	19
# NIL Person	150	76
# NIL Product	43	12
# Thing	6	0
# Event	6	12
# Character	12	2
# Location	116	70
# Organization	148	56
# Person	173	61
# Product	65	25
# NIL instances	275	131
# Entities	526	357

for all annotations considering the mention boundaries and their types. This is a measure of the tagging capability of the system.

SLM (*Strong Link Match*). This metrics is the micro average F-1 score for annotations considering the correct link for each mention. This is a measure of the linking performance of the system.

MC (*Mention Ceaf*). This metrics, also known as Constrained Entity-Alignment F-measure (Luo, 2005), is a clustering metric developed to evaluate clusters of annotations. It evaluates the F-1 score for both NIL and non-NIL annotations in a set of mentions.

The final score for each system is a combination of the aforementioned metrics and is computed as follows:

$$score = 0.4 \times MC + 0.3 \times STMM + 0.3 \times SLM. \quad (1)$$

All the metrics were computed by using the TAC KBP scorer⁴.

⁴<https://github.com/wikilinks/neleval/>

5 Systems Description

The task was well received by the NLP community and was able to attract 17 participants who expressed their interest in the evaluation. Five groups participated actively to the challenge by submitting their system results, each group presented three different runs, for a total amount of 15 runs submitted. In this section we briefly describe the methodology followed by each group.

5.1 UniPI

The system proposed by the University of Pisa (Attardi et al., 2016) exploits word embeddings and a bidirectional LSTM for entity recognition and linking. The team produced also a training dataset of about 13,945 tweets for entity recognition by exploiting active learning, training data taken from the PoSTWITA task (Tamburini et al., 2016) and manual annotation. This resource, in addition to word embeddings built on a large corpus of Italian tweets, is used to train a bidirectional LSTM for the entity recognition step. In the linking step, for each Wikipedia page its abstract is extracted and the average of the word embeddings is computed. For each candidate entity in the tweet, the word embedding for a context of words of size c before and after the entity is created. The linking is performed by comparing the mention embedding with the DBpedia entity whose lc2 distance is the smallest among those entities whose abstract embeddings were computed at the previous step. The Twitter mentions were resolved by retrieving the real name with the Twitter API and looking up in a gazetteer in order to identify the Person-type entities.

5.2 MicroNeel

MicroNeel (Corcoglioniti et al., 2016) investigates the use on microposts of two standard NER and Entity Linking tools originally developed for more formal texts, namely Tint (Palmero Arosio and Moretti, 2016) and The Wiki Machine (Palmero Arosio and Giuliano, 2016). Comprehensive tweet preprocessing is performed to reduce noisiness and increase textual context. Existing alignments between Twitter user profiles and DBpedia entities from the Social Media Toolkit (Nechaev et al., 2016) resource are exploited to annotate user mentions in the tweets.

Rule-based and supervised (SVM-based) techniques are investigated to merge annotations from different tools and solve possible conflicts. All the resources listed as follows were employed in the evaluation:

- The Wiki Machine (Palmero Arosio and Giuliano, 2016): an open source entity linking for Wikipedia and multiple languages.
- Tint (Palmero Arosio and Moretti, 2016): an open source suite of NLP modules for Italian, based on Stanford CoreNLP, which supports named entity recognition.
- Social Media Toolkit (SMT) (Nechaev et al., 2016): a resource and API supporting the alignment of Twitter user profiles to the corresponding DBpedia entities.
- Twitter ReST API⁵: a public API for retrieving Twitter user profiles and tweet metadata.
- Morph-It! (Zanchetta and Baroni, 2005): a free morphological resource for Italian used for preprocessing (true-casing) and as source of features for the supervised merging of annotations.
- tagdef⁶: a website collecting user-contributed descriptions of hashtags.
- list of slang terms from Wikipedia⁷.

5.3 FBK-HLT-NLP

The system proposed by the FBK-HLT-NLP team (Minard et al., 2016a) follows 3 steps: entity recognition and classification, entity linking to DBpedia and clustering. Entity recognition and classification is performed by the EntityPro module (included in the TextPro pipeline), which is based on machine learning and uses the SVM algorithm. Entity linking is performed using the named entity disambiguation module developed within the NewsReader and based on DBpedia Spotlight. The FBK team exploited a specific resource to link the Twitter profiles to DBpedia: the Alignments dataset. The clustering step is string-based, i.e. two entities are part of the same cluster if they are equal.

⁵<https://dev.twitter.com/rest/public>

⁶<https://www.tagdef.com/>

⁷https://it.wikipedia.org/wiki/Gergo_di_Internet

Moreover, the FBK team exploits active learning for domain adaptation, in particular to adapt a general purpose Named Entity Recognition system to a specific domain (tweets) by creating new annotated data. In total they have annotated 2,654 tweets.

5.4 Sisinflab

The system proposed by Sisinflab (Cozza et al., 2016) faces the neel-it challenge through an ensemble approach that combines unsupervised and supervised methods. The system merges results achieved by three strategies:

1. DBpedia Spotlight for span and URI detection plus SPARQL queries to DBpedia for type detection;
2. Stanford CRF-NER trained with the challenge train corpus for span and type detection and DBpedia lookup for URI detection;
3. DeepNL-NER, a deep learning classifier trained with the challenge train corpus for span and type detection, it exploits ad-hoc gazetteers and word embedding vectors computed with word2vec trained over the Twita dataset⁸ (a subset of 12,000,000 tweets). DBpedia is used for URI detection.

Finally, the system computes NIL clusters for those mentions that do not match with an entry in DBpedia, by grouping in the same cluster entities with the same text (no matter the case). The Sisinflab team submitted three runs combining the previous strategies, in particular: run1 combines (1), (2) and (3); run2 involves strategies (1) and (3); run3 exploits strategies (1) and (2).

5.5 UNIMIB

The system proposed by the UNIMIB team (Cecchini et al., 2016) is composed of three steps: 1) Named Entity Recognition using Conditional Random Fields (CRF); 2) Named Entity Linking by considering both Supervised and Neural-Network Language models and 3) NIL clustering by using a graph-based approach. In the first step two kinds of CRF are exploited: 1) a simple CRF on the training data and 2) CRF+Gazetteers, in this

⁸<http://www.let.rug.nl/basile/files/proc/>

configuration the model has been induced by exploiting several gazetteers, i.e. products, organizations, persons, events and characters. Two strategies are adopted for the linking. A decision strategy is used to select the best link by exploiting a large set of supervised methods. Then, word embeddings built on Wikipedia are used to compute a similarity measure used to select the best link for a list of candidate entities. NIL clustering is performed by a graph-based approach; in particular, a weighted indirect co-occurrence graph where an edge represents the co-occurrence of two terms in a tweet is built. The ensuing word graph was then clustered using the MaxMax algorithm.

6 Results

The performance of the participant systems were assessed by exploiting the final score measure presented in Eq. 1. This measure combines the three different aspects evaluated during the task, i.e. the correct tagging of the mentions (STMM), the proper linking to the knowledge base (SLM), and the clustering of the NIL instances (MC). Results of the evaluation in terms of the final score are reported in Table 3.

The best result was reported by *Uni.PI.3*, this system obtained the best final score of 0.5034 with an improvement with respect to the *Uni.PI.1* (second classified) of +1.27. The difference between these two runs lays on the different vector dimension (200 in *Uni.PI.3* rather than 100 in *Uni.PI.1*) combined with the use of Wikipedia embeddings and a specific training set for geographical entities (*Uni.PI.3*) rather than a mention frequency strategy for disambiguation (*Uni.PI.1*). *MicroNeel.base* and *FBK-HLT-NLP* obtain remarkable results very close to the best system. Indeed, *MicroNeel.base* reported the highest linking performance (SLM = 0.477) while *FBK-HLT-NLP* showed the best clustering (MC = 0.585) and tagging (STMM = 0.516) results. It is interesting to notice that all these systems (*Uni.PI*, *MicroNeel* and *FBK-HLT-NLP*) developed specific techniques for dealing with Twitter mentions reporting very good results for the tagging metric (with values always above 0.46).

All participants have made use of supervised algorithms at some point of their tagging/linking/clustering pipeline. *Uni.PI*, *Sisinflab* and *UNIMIB* have exploited word embeddings trained on the development set plus some

other external resources (manual annotated corpus, Wikipedia, and Twita). *UniPI* and *FBK-HLT-NLP* built additional training data obtained by active learning and manual annotation. The use of additional resources is allowed by the task guidelines, and both the teams have contributed to develop additional data useful for the research community.

7 Conclusions

We described the first evaluation task for entity linking in Italian tweets. The task evaluated the performance of participant systems in terms of (1) tagging entity mentions in the text of tweets; (2) linking the mentions with respect to the canonicalized DBpedia 2015-10; (3) clustering the entity mentions that refer to the same named entity.

The task has attracted many participants who specifically designed and developed algorithm for dealing with both Italian language and the specific peculiarity of text on Twitter. Indeed, many participants developed ad-hoc techniques for recognising Twitter mentions and hashtag. In addition, the participation in the task has fostered the building of new annotated datasets and corpora for the purpose of training learning algorithms and word embeddings.

We hope that this first initiative has set up the scene for further investigations and developments of best practises, corpora and resources for the Italian name entity linking on Tweets and other microblog contents.

As future work, we plan to build a bigger dataset of annotated contents and to foster the release of state-of-the-art methods for entity linking in Italian language.

Acknowledgments

This work is supported by the project “Multilingual Entity Linking” co-funded by the Apulia Region under the program FutureInResearch, by the ADAPT Centre for Digital Content Technology, which is funded under the Science Foundation Ireland Research Centres Programme (Grant 13/RC/2106) and is co-funded under the European Regional Development Fund, and by H2020 FREME project (GA no. 644771).

References

- Giuseppe Attardi, Daniele Sartiano, Maria Simi, and Irene Sucameli. 2016. Using Embeddings for Both Entity Recognition and Linking in Tweets. In Pierpaolo Basile, Franco Cutugno, Malvina Nissim, Viviana Patti, and Rachele Sprugnoli, editors, *Proceedings of the 5th Evaluation Campaign of Natural Language Processing and Speech Tools for Italian (EVALITA 2016)*. aAcademia University Press.
- Timothy Baldwin, Young-Bum Kim, Marie Catherine de Marneffe, Alan Ritter, Bo Han, and Wei Xu. 2015. Shared tasks of the 2015 workshop on noisy user-generated text: Twitter lexical normalization and named entity recognition. *ACL-IJCNLP*, 126:2015.
- Francesco Barbieri, Valerio Basile, Danilo Croce, Malvina Nissim, Nicole Novielli, and Viviana Patti. 2016. Overview of the EVALITA 2016 SENTiment POLarity Classification Task. In Pierpaolo Basile, Franco Cutugno, Malvina Nissim, Viviana Patti, and Rachele Sprugnoli, editors, *Proceedings of the 5th Evaluation Campaign of Natural Language Processing and Speech Tools for Italian (EVALITA 2016)*. aAcademia University Press.
- Valerio Basile and Malvina Nissim. 2013. Sentiment analysis on italian tweets. In *Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 100–107, Atlanta, Georgia, June. Association for Computational Linguistics.
- Pierpaolo Basile, Annalina Caputo, and Giovanni Semeraro. 2015. Entity Linking for Italian Tweets. In Cristina Bosco, Sara Tonelli, and Fabio Massimo Zanzotto, editors, *Proceedings of the Second Italian Conference on Computational Linguistics CLiC-it 2015, Trento, Italy, December 3-8, 2015.*, pages 36–40. Accademia University Press.
- Flavio Massimiliano Cecchini, Elisabetta Fersini, Enza Messina Pikakshi Manchanda, Debora Nozza, Matteo Palmonari, and Cezar Sas. 2016. UNIMIB@NEEL-IT : Named Entity Recognition and Linking of Italian Tweets. In Pierpaolo Basile, Franco Cutugno, Malvina Nissim, Viviana Patti, and Rachele Sprugnoli, editors, *Proceedings of the 5th Evaluation Campaign of Natural Language Processing and Speech Tools for Italian (EVALITA 2016)*. aAcademia University Press.
- Francesco Corcoglioniti, Alessio Palmero Aprosio, Yaroslav Nechaev, and Claudio Giuliano. 2016. MicroNeel: Combining NLP Tools to Perform Named Entity Detection and Linking on Microposts. In Pierpaolo Basile, Franco Cutugno, Malvina Nissim, Viviana Patti, and Rachele Sprugnoli, editors, *Proceedings of the 5th Evaluation Campaign of Natural Language Processing and Speech Tools for Italian (EVALITA 2016)*. aAcademia University Press.
- Vittoria Cozza, Wanda La Bruna, and Tommaso Di Noia. 2016. sisinflab: an ensemble of supervised

Table 3: Results of the evaluation with respect to: MC (*Mention_Ceaf*), STMM (*Strong_Typed_Mention_Match*), SLM (*Strong_Link_Match*) and the final score used for system ranking. Δ shows the final score improvement of the current system versus the previous. Best MC, STMM and SLM are reported in bold.

name	MC	STMM	SLM	final score	Δ
UniPI.3	0.561	0.474	0.456	0.5034	+1.27
UniPI.1	0.561	0.466	0.443	0.4971	+0.08
MicroNeel.base	0.530	0.472	0.477	0.4967	+0.10
UniPI.2	0.561	0.463	0.443	0.4962	+0.61
FBK-HLT-NLP.3	0.585	0.516	0.348	0.4932	+0.78
FBK-HLT-NLP.2	0.583	0.508	0.346	0.4894	+1.49
FBK-HLT-NLP.1	0.574	0.509	0.333	0.4822	+1.49
MicroNeel.merger	0.509	0.463	0.442	0.4751	+0.32
MicroNeel.all	0.506	0.460	0.444	0.4736	+38.56
sisinflab.1	0.358	0.282	0.380	0.3418	0.00
sisinflab.3	0.358	0.286	0.376	0.3418	+2.24
sisinflab.2	0.340	0.280	0.381	0.3343	+50.31
unimib.run_02	0.208	0.194	0.270	0.2224	+9.50
unimib.run_03	0.207	0.188	0.213	0.2031	+5.56
unimib.run_01	0.193	0.166	0.218	0.1924	0.00

- and unsupervised strategies for the neel-it challenge at Evalita 2016. In Pierpaolo Basile, Franco Cutugno, Malvina Nissim, Viviana Patti, and Rachele Sprugnoli, editors, *Proceedings of the 5th Evaluation Campaign of Natural Language Processing and Speech Tools for Italian (EVALITA 2016)*. aAcademia University Press.
- Aba-Sah Dadzie, Daniel PreoĂciuc-Pietro, Danica RadovanoviĂG, Amparo E. Cano Basave, and Katrin Weller, editors. 2016. *Proceedings of the 6th Workshop on Making Sense of Microposts*, volume 1691. CEUR.
- Xiaoqiang Luo. 2005. On coreference resolution performance metrics. In *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, pages 25–32. Association for Computational Linguistics.
- Anne-Lyse Minard, R. H. Mohammed Qwaider, and Bernardo Magnini. 2016a. FBK-NLP at NEEL-IT: Active Learning for Domain Adaptation. In Pierpaolo Basile, Franco Cutugno, Malvina Nissim, Viviana Patti, and Rachele Sprugnoli, editors, *Proceedings of the 5th Evaluation Campaign of Natural Language Processing and Speech Tools for Italian (EVALITA 2016)*. aAcademia University Press.
- Anne-Lyse Minard, Manuela Speranza, and Tommaso Caselli. 2016b. The EVALITA 2016 Event Factuality Annotation Task (FactA). In Pierpaolo Basile, Franco Cutugno, Malvina Nissim, Viviana Patti, and Rachele Sprugnoli, editors, *Proceedings of the 5th Evaluation Campaign of Natural Language Processing and Speech Tools for Italian (EVALITA 2016)*. aAcademia University Press.
- Yaroslav Nechaev, Francesco Corcoglioniti, and Claudio Giuliano. 2016. Linking knowledge bases to social media profiles.
- Alessio Palmero Aprosio and Claudio Giuliano. 2016. The Wiki Machine: an open source software for entity linking and enrichment. *ArXiv e-prints*.
- Alessio Palmero Aprosio and Giovanni Moretti. 2016. Italy goes to Stanford: a collection of CoreNLP modules for Italian. *ArXiv e-prints*, September.
- Giuseppe Rizzo, Marieke van Erp, Julien Plu, and Raphaël Troncy. 2016. Making Sense of Microposts (#Microposts2016) Named Entity rEcognition and Linking (NEEL) Challenge. In *6th Workshop on Making Sense of Microposts (#Microposts2016)*.
- Pontus Stenetorp, Sampo Pyysalo, Goran Topić, Tomoko Ohta, Sophia Ananiadou, and Jun’ichi Tsujii. 2012. Brat: A web-based tool for nlp-assisted text annotation. In *Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics, EACL ’12*, pages 102–107, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Fabio Tamburini, Cristina Bosco, Alessandro Mazzei, and Andrea Bolioli. 2016. Overview of the EVALITA 2016 Part Of Speech on TWitter for ITALian Task. In Pierpaolo Basile, Franco Cutugno, Malvina Nissim, Viviana Patti, and Rachele Sprugnoli, editors, *Proceedings of the 5th Evaluation Campaign of Natural Language Processing*

and Speech Tools for Italian (EVALITA 2016). aAcademia University Press.

Eros Zanchetta and Marco Baroni. 2005. Morph-it! a free corpus-based morphological resource for the Italian language. *Corpus Linguistics 2005*, 1(1).