

# Transformer-based Methods for Recognizing Ultra Fine-grained Entities (RUFES)

Emanuela Boros and Antoine Doucet

L3i, University of La Rochelle,

F-17000, La Rochelle, France

{emanuela.boros, antoine.doucet}@univ-lr.fr

## Abstract

This paper summarizes the participation of the *Laboratoire Informatique, Image et Interaction* (L3i laboratory) of the University of La Rochelle in the Recognizing Ultra Fine-grained Entities (RUFES) track<sup>1</sup> within the Text Analysis Conference (TAC) series of evaluation workshops. Our participation relies on two neural-based models, one based on a pre-trained and fine-tuned language model with a stack of Transformer layers for fine-grained entity extraction and one out-of-the-box model for within-document entity coreference. We observe that our approach has great potential in increasing the performance of fine-grained entity recognition. Thus, the future work envisioned is to enhance the ability of the models following additional experiments and a deeper analysis of the results.

## 1 Introduction

Fine-grained entity recognition aims at labeling entity mentions in context with one or more specific types organized in a hierarchy (e.g., *Photographer* is from a *Artist* that in turn, is a subtype of *PER*<sup>2</sup>). The need for a wider variety of fine-grained entities (e.g., technical terms, lawsuits, disease, crisis, biomedical entities) (Ji et al., 2019) can support the development of real-world applications that combine several information sources that include both text sources and knowledge bases; e.g., question answering (Lin and Ji, 2019), relation extraction systems (Yao et al., 2010) that need access to knowledge bases i.e. gazetteers, or named entity recognition systems (Ehrmann et al., 2020a,b) for supporting more accurate entity linking in historical documents.

In the context of the Recognizing Ultra Fine-grained Entities (RUFES) track<sup>3</sup> of the Text Anal-

ysis Conference (TAC) series of evaluation workshops, the main task consisted in automatically identifying fine-grained entities as clusters of names, nominals, and/or pronominal mentions, and classifying them into one or more of the types defined in a detailed ontology developed by NIST<sup>4</sup>. This year, the track focused on document-level entity discovery and only on English source documents.

The track had two phases: a preliminary phase where the data is provided along with a limited annotated set of samples (50 documents), and a second phase during which human feedback was provided for the preliminary submissions based on a user model of how analysts might interact with the systems. The final results include this feedback.

This paper presents the participation of the *Laboratoire Informatique, Image et Interaction* (L3i laboratory) at the University of La Rochelle at TAC KBP RUFES 2020. We applied our recent proposed model for coarse-grained and fine-grained named entity recognition (Boros et al., 2020a,b) and we used an out-of-the-box neural-based entity coreference model for detecting the mentions that refer to the same entity.

The remaining of the paper is organized as follows: we detail the data pre-processing step and our proposed methods in Section 2. Section 3 presents the results and Section 4 finalizes the paper with several conclusions and envisioned future work.

## 2 Methods

We separated RUFES in two sub-tasks:

- Entity extraction: the detection and the classification of fine-grained entity types including the named, nominal, and pronominal mentions for each mention (labeled as NAM, NOM, and PRO, respectively);

<sup>1</sup><https://tac.nist.gov/2020/KBP/RUFES/>

<sup>2</sup>*PER* refers to the entity type *Person*.

<sup>3</sup><https://tac.nist.gov/2020/KBP/RUFES/>

<sup>4</sup><https://www.nist.gov/>

- Within-document entity coreference resolution: the detection of the referential mentions in a document that point to the same entity.

## 2.1 Data Pre-processing

The KBP 2020 RUFES dataset provided by the organizers consisted of the development source documents and evaluation source documents drawn from a collection of Washington Post news articles.

The development source corpus and the evaluation source corpus each comprised approximately 100,000 articles from which 50 documents were annotated for the development set with entity types from an ontology that contains approximately 200 fine-grained entity types and that followed the same three-level x.y.z hierarchy as in the TAC-KBP 2019 EDL track (Ji et al., 2019).

The provided data was organized into two formats: `./rsd/`: “raw source data” (rsd) plain text form of the new article; and `./ltf/`: “logical text format” (ltf) derived from the rsd version.

For the data pre-processing, we used the `ltf.xml` files that each contained a fully segmented and tokenized version of the corresponding rsd file<sup>5</sup>. Next, we converted the data in IOB<sup>6</sup> format as shown in the following example:

TOKEN	FIRST_LEVEL	SECOND_LEVEL	THIRD_LEVEL	FORTH_LEVEL
Georgetown	B-ORG	B-EducationalInstitution	B-College	B-NAM
University	I-ORG	I-EducationalInstitution	I-College	I-NAM
officials	B-PER	B-Professional	O	B-NOM
on	O	O	O	O
Thursday	O	O	O	O
announced	O	O	O	O
that	O	O	O	O
they	B-ORG	B-EducationalInstitution	B-College	B-NOM
would	O	O	O	O
build	O	O	O	O
additional	O	O	O	O
on	O	O	O	O
-	O	O	O	O
campus	O	O	O	O
student	O	O	O	O
housing	B-FAC	B-Building	B-ApartmentBuilding	B-NOM

Figure 1: Data formatting example for the KBP 2020 RUFES dataset.

## 2.2 Entity Extraction

Due to the complexity and characteristics of the fine-grained sub-task, we made use of our recently proposed model for coarse-grained and fine-grained named entity recognition (Boros et al., 2020a,b) that consists in a hierarchical, multitask learning approach, with a fine-tuned encoder based

<sup>5</sup>Segments (paragraphs) and the tokens (words) are marked off by XML tags (SEG and TOKEN), with “id” attributes (which are only unique within a given XML file) and character offset attributes relative to the corresponding `rsd.txt` file.

<sup>6</sup>[https://en.wikipedia.org/wiki/Inside%20%93outside%20%93beginning\\_\(tagging\)](https://en.wikipedia.org/wiki/Inside%20%93outside%20%93beginning_(tagging))

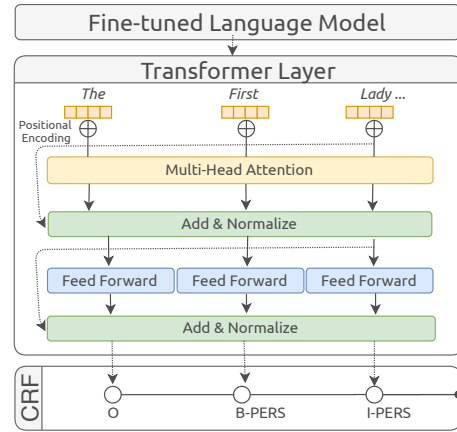


Figure 2: Detailed model proposed for entity extraction in Boros et al. (2020a,b).

on *Bidirectional Encoder Representations from Transformers* (BERT) (Devlin et al., 2019).

This model includes the use of a stack of Transformer (Vaswani et al., 2017) blocks on top of the BERT encoder. The multitask prediction layer consists of separate conditional random field (CRF) layers. The architecture of the model is presented in Figure 2.

The added layers are a stack of Transformer blocks (encoders). As proposed by Vaswani et al. (2017), this model is a deep learning architecture based on multi-head attention mechanisms with *sinusoidal position embeddings*<sup>7</sup>. It is composed of a stack of identical layers. Each layer has two sub-layers. The first layer is a multi-head self-attention mechanism, while the second one is a simple, position-wise fully connected feed-forward network. A residual connection is around each of the two sub-layers, followed by layer normalization.

We decided to add a stack of Transformer layers due to the assumption that additional hyperparameters can increase the ability of the architecture to better model long-range contexts and alleviate the number of spurious predicted entities, as observed in Boros et al. (2020a).

Because of multitask learning, this method has a label independence assumption, which is not valid for fine-grained entity typing. For example, if the model is confident at predicting the type *Photographer*, it should promote its parents of type *Artist* and *PER*, but discourage entity types as in *ORG*

<sup>7</sup>In our implementation, we used *learned absolute positional embeddings* (Gehring et al., 2017) instead, as suggested by Wolf et al. (2019). Vaswani et al. (2017) found that both versions produced nearly identical results.

and its descendant types (i.e. *Association*, *CommercialOrganization*). In order to capture interdependencies between types, we post-process the predictions by checking them again against the ontology terms, and by offering more confidence to the last predicted entity subtype. For example, if we take the subtype *ProvinceState* with its parent *GPE*, and if “Illinois” was recognized as *LOC.ProvinceState*, we lookup *ProvinceState* in the ontology and we automatically correct the prediction to *GPE.ProvinceState*.

**Parameters** For the pre-trained BERT encoder, we used the `bert-large-cased` model. We added two Transformer layers, with the hidden size of 128 and 12 number of self-attention heads<sup>8</sup>. We trained for 10 epochs, with Adam optimizer with weight decay,  $2 \times 10^{-5}$  learning rate and a mini-batch of dimension 8.

### 2.3 Within-document Entity Coreference Resolution

For detecting the referential mentions in a document that point to the same entity, we used an out-of-the-box tool, NeuralCoref<sup>9</sup>, which is a pipeline extension for spaCy 2.1+<sup>10</sup> (Honnibal and Montani, 2017) that annotates and resolves coreference clusters using a neural-based method. Due to a lack of time and resources, we did not re-train this model on the KBP 2020 RUFES dataset. The model was previously trained on OntoNotes 5.0 dataset<sup>11</sup>.

NeuralCoref has two sub-modules:

- a rule-based mentions detection module which uses spaCy’s tagger, parser and entity annotations to identify a set of potential coreference mentions;
- a feed-forward neural-network which compute a coreference score for each pair of potential mentions. This scoring system is an adaptation of Clark and Manning (2016a,b).

We applied this model in a within-document context, with the default parameters.

<sup>8</sup>The parameters correspond to the best configuration reported by Boros et al. (2020a).

<sup>9</sup><https://github.com/huggingface/neuralcoref>

<sup>10</sup><https://spacy.io/>

<sup>11</sup>[https://www.gabormelli.com/RKB/OntoNotes\\_Corpus](https://www.gabormelli.com/RKB/OntoNotes_Corpus)

### 2.4 Rule-based Feedback Inclusion

In the second phase of the track, after receiving the feedback, we inspected the most frequent types of error produced by our methods.

A majority of the mistakes, around 46% out of 400 (the first 40 errors detected in ten random documents were reported) were system mention-level entity types that do not exactly match the gold mention-level entity types (including the level of granularity). Around 11% were extraneous mentions (a system mention span does not exactly match or overlap with any gold mention span), 5% of wrong extents (a system mention span and gold mention span overlap but have different extents), 12% missing mentions (a gold mention span does not exactly match any system mention span), and around 11% had the wrong entity coreference, either missing, incorrect or spurious (problems in linking a system mention to a mention of the same entity that occurred earlier in the document).

Thus, we focused on the feedback related to the detection of the wrong type of entity (46% out of all errors), and we integrated this feedback in a rule-based manner by automatically creating a set of rules to change the predictions accordingly. One common mistake produced by our system was related to entities that had one of the ontology terms included in the entity; For instance, “Norovirus” was recognized as GPE (geopolitical entity) instead of Pathogen.Virus. Therefore, for every entity that included a fine-grained ontology type i.e. “Airport”, “Hospital”, “Highway”, a rule was created to change the predictions into the correct types.

## 3 Results

In the initial phase, we submitted two runs named *1-first-rufes* and *2-first-rufes*. In the second phase, after receiving feedback for a sample of our previous predictions, we submitted another two runs referred to as *1-feedback-rufes* and *2-feedback-rufes*.

The metrics reported in Table 1 are implemented by `neleval`<sup>12</sup>. For the entity extraction evaluation, we report the *strong\_mention\_match* (where an entity span must match a gold span exactly to be counted as correct) and the *strong\_typed\_mention\_match* (that additionally requires the correct entity type). For the entity coreference evaluation (Pradhan et al., 2014), we report

<sup>12</sup>More details about the TAC evaluation tool can be found here <https://neleval.readthedocs.io/en/latest/>.

Submission	strong mention match	strong typed mention match	mention ceaf	typed mention ceaf	entity ceaf	fine grain typing
1-first-rufes	<b>0.868</b>	0.745	0.552	0.503	0.551	0.3188
2-first-rufes	<b>0.868</b>	0.745	0.578	0.503	0.567	0.3188
1-feedback-rufes	<b>0.868</b>	0.745	0.578	0.504	0.567	0.3204
2-feedback-rufes	<b>0.868</b>	0.745	0.578	0.504	0.567	0.3239
Median	0.805	–	–	–	0.578	0.2313
Maximum	<b>0.868</b>	–	–	–	0.689	<b>0.4162</b>

Table 1: The scoring results (F-score) for RUFES 2020 evaluation for all our submissions. Median and Maximum scores are computed on the best-performing submission from each participant, as shared by RUFES organizers.

the *typed\_mention\_ceaf*, the *entity\_ceaf*, and *mention\_ceaf* metrics. We also report the RUFES final score, *fine\_grain\_typing*.

From the scoring results for our submissions reported in the Table 1, we notice that the impact of the ruled-based feedback inclusion is rather insignificant. The slight differences in the scores when comparing *1-first-rufes* and the other submissions regarding the *entity\_ceaf* and *mention\_ceaf* metrics are due to the fact that we did not approach the entity coreference for our first submission.

## 4 Conclusions

This paper described our approach for the 2020 TAC RUFES task that implied fine-grained entity recognition and within-document entity coreference. We presented our proposed models, and we reported the results obtained in the context of the track. In future work, we will focus on enhancing the ability of the models following additional experiments by refining the entity extraction architecture in order to be able to take into consideration the inter-dependencies between entity types. Regarding the entity coreference model, we could explore a further fine-tuning of the out-of-the-box model on the KBP RUFES annotated documents. Moreover, a deeper and more qualitative analysis of the types of errors is intended.

## Acknowledgments

This work has been supported by the European Union’s Horizon 2020 research and innovation program under grants 770299 (NewsEye) and 825153 (Embeddia).

## References

- Emanuela Boros, Ahmed Hamdi, Elvys Linhares Pontes, Luis-Adrián Cabrera-Diego, José G Moreno, Nicolas Sidere, and Antoine Doucet. 2020a. Alleviating digitization errors in named entity recognition for historical documents. In *Proceedings of the 24th Conference on Computational Natural Language Learning*, pages 431–441.
- Emanuela Boros, Elvys Linhares Pontes, Luis Adrián Cabrera-Diego, Ahmed Hamdi, José Moreno, Nicolas Sidère, and Antoine Doucet. 2020b. Robust named entity recognition and linking on historical multilingual documents. In *Conference and Labs of the Evaluation Forum (CLEF 2020)*, volume 2696, pages 1–17. CEUR-WS Working Notes.
- Kevin Clark and Christopher D Manning. 2016a. Deep reinforcement learning for mention-ranking coreference models. *arXiv preprint arXiv:1609.08667*.
- Kevin Clark and Christopher D Manning. 2016b. Improving coreference resolution by learning entity-level distributed representations. *arXiv preprint arXiv:1606.01323*.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics*, pages 4171–4186.
- Maud Ehrmann, Matteo Romanello, Stefan Bircher, and Simon Clematide. 2020a. Introducing the CLEF 2020 HIPE shared task: Named entity recognition and linking on historical newspapers. In *Advances in information retrieval*, pages 524–532, Cham. Springer International Publishing.
- Maud Ehrmann, Matteo Romanello, Alex Flückiger, and Simon Clematide. 2020b. Impresso named entity annotation guidelines (version 2.2.0). <https://doi.org/10.5281/zenodo.3604227>.
- Jonas Gehring, Michael Auli, David Grangier, Denis Yarats, and Yann N Dauphin. 2017. Convolutional neural networks for machine translation.

tional sequence to sequence learning. *arXiv preprint arXiv:1705.03122*.

Matthew Honnibal and Ines Montani. 2017. spacy 2: Natural language understanding with bloom embeddings, convolutional neural networks and incremental parsing.

Heng Ji, Avirup Sil, Hoa Trang Dang, Ian Soboroff, Joel Nothman, and Sydney Informatics Hub. 2019. Overview of tac-kbp2019 fine-grained entity extraction. In *2019 Text Analysis Conference Proceedings*.

Ying Lin and Heng Ji. 2019. An attentive fine-grained entity typing model with latent type representation. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 6198–6203.

Sameer Pradhan, Xiaoqiang Luo, Marta Recasens, Eduard Hovy, Vincent Ng, and Michael Strube. 2014. Scoring coreference partitions of predicted mentions: A reference implementation. In *Proceedings of the conference. Association for Computational Linguistics. Meeting*, volume 2014, page 30. NIH Public Access.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in neural information processing systems*, pages 5998–6008.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, R’emi Louf, Morgan Funtowicz, and Jamie Brew. 2019. Huggingface’s transformers: State-of-the-art natural language processing. *ArXiv*, abs/1910.03771.

Limin Yao, Sebastian Riedel, and Andrew McCallum. 2010. Collective cross-document relation extraction without labelled data. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pages 1013–1023.