IXA-NER-RE at eHealth-KD Challenge 2020: Cross-Lingual Transfer Learning for Medical Relation Extraction

Edgar Andrés[a], Oscar Sainz[a], Aitziber Atutxa[a] and Oier Lopez de Lacalle[a]

[a]*IXA NLP Group, University of the Basque Country (UPV/EHU)*

Abstract

The eHealth-KD 2020 set out this year an automatic extraction challenge on a coarse range of knowledge from health documents written in the Spanish Language. Our group has participated in all the proposed scenarios; the main one, the Named Entity Recognition (NER) subtask, the Relation Extraction (RE) subtask, and the alternative domain obtaining very different results in each of them. The main task has been conceived as a pipeline of the NER and RE subtask, each of them independently developed from the other. The Name Entity Recognition task has been envisaged as a basic seq2seq system applying a general-purpose Language Model and static embeddings. Unlike the NER subtask, in the RE subtask several approaches were successfully explored; first, transfer learning methods as a way to measure the adaptation ability of pre-trained language models to both medical domain and Spanish language. Second, Matching the Blanks to tackle the problem of the reduced size of the training corpus by producing relation representations directly from non tagged text. As mentioned, the results in the different task were heterogeneous; while the result in NER is on the average (F1 0.66), with ample room for improvement, the result in RE has been outstanding, obtaining the first place in this task (F1 0.633) with more than 3 points over the next classified, demonstrating the soundness of the proposed techniques.

Keywords

Language Models, Matching the Blanks, Named Entity Recognition, Relation Extraction

1. Introduction

In this paper we describe our participation at eHealth-KD 2020 shared task [1], consisting on extracting semantic structured information for Spanish medical texts. The challenge is divided in two main tasks proposed as a pipeline. The first task is devoted to the identification and classification of medical entities. In the second task, participants need to detect the semantic relations between the entities, presumably, discovered in the first task.

Organizers proposed different evaluation schemes in which 1) systems are evaluated on the whole tasks at once (*main evaluation*), and 2) entity recognition and relation extraction are evaluated separately (*task A* and *task B*, respectively). Our system is built on top of two independent components and, thus, training and development of the each component is carried out in their specific sub-tasks separately.

We approached the Named Entity Recognition (NER) task[1] with a character based BiLSTM sequence labeler [2] trained on the training set provided by the organizers employing both pre-trained general domain Language Model and static word embeddings. Regarding the relation extraction task[2], in order to solve it, we decided to use a transfer learning strategy and fine-tune existing multilingual pre-trained language models [3] in the annotated data of the task.

Note that we propose a system with heterogeneous components, with different goals for each of the part in the system. This way, the goals of our participation in the task are twofold.

- **Entity recognition**: Our goal was to check the suitability of a character based pre-trained Language Model based system on a heterogeneous NER setting. Pre-trained Language Model based systems have been successfully used in Medical Entity Recognition (MER) tasks. But unlike other similar challenges that involved MER (CLEF eHealth 2020, PharmacoNER 2019), the present task is especially challenging because the entities are not purely medical but very heterogeneous not only semantically regarding the domain, but also syntactically.

- **Relation extraction**: Our main goal for using large multilingual pre-trained language models is to measure ability to adapt to medical domain and Spanish language when using transfer learning methods. In addition make experiments adapting Matching the Blanks [4] (MTB) to eHealth-KD 2020 setting.

The system obtains very uncompensated results, while in relation extraction we outperform the rest of the participants by wide margin (3.4 points better than the second ranked), system has large room of improvement in entity recognition (we are still more than 10 points lower than best systems). Overall, our system shows very competitive results with 55.7 of F1 in the main task.

2. Related Work

Entity recognition   MER, as opposed to NER, shows certain specificities [5], like their descriptive nature, their productivity and the massive use of acronyms. These specificities and the fact that static embeddings were systematically employed by NER systems, yield researchers to use in-domain corpus, as opposed to general-domain corpus, to both train the MER systems as well as the static pre-trained embeddings, in order to obtain better results since controlling domain leads to better control on polysemy ([6],[7]). Recently, performance of both NER and MER tasks has shown a significant breakthrough with the introduction of contextualized word embeddings (ELMo [8], ULMFiT [9], BERT [10] and FLAIR [2]).

Although contextualized embeddings seem to reduce the gap between general and domain specific corpus, several works on MER task argue that domain-specific contextualized embeddings still yields superior performance over the standard and general-domain word embeddings ([11],[12],[13],[14]). As mentioned in the introduction, the present MER task due to its heterogeneity (`concepts` are more specific to the medical domain while `actions` or `references`

---

[1]Entities are classified in 4 types: `concept`, `action`, `predicate`, and `reference`.

[2]The 13 relation types are organized in 4 main categories: `general relations`, `contextual relations`, `action roles`, and `predicate roles`.

are less specific) represents a perfect playground to check the performance of contextualized Language Models based embeddings calculated over general domain corpus on the different entities.

Transfer Learning   Recently transfer learning has been shown as a successful alternative when (almost) no annotated data is available in the target domain and language [10, 4]. Recent Transformer sequence models [15] surpass the state-of-the-art in many information extraction tasks such as relation extraction [4, 16, 17]. Some works try to integrate the information available in knowledge bases into Transformers sequence models [16]. Nevertheless, simpler approaches based on entity markers (further details in Section 4) show same competitive performance with a quicker setup [4]. In a similar manner, multilingual language models [3] have shown impressive capacity to perform zero-shot learning in cross-lingual tasks. This kind of models seems very promising for relation extraction tasks where target language contains small annotated training set.

Data-augmentation   A variety of data-augmentation have been proposed for information extraction tasks. One of the most significant paradigm is distant-supervision [18, 19], in which existing relations in knowledge bases are aligned to unlabeled text relying on some heuristics and automatically labeling training data [20]. More recently, Soares et al. [4] introduce an augmentation method that does not require relation label and adapts the model learning by *Matching the Blanks* (MTB). In this work we explore the idea of MTB to approach the relation extraction task in eHealth-KD 2020.

3. Entity Recognition system

We adopted a sequence to sequence Deep Learning approach [2] to pursue the Named Entity Recognition (NER) task.

3.1. NER Architecture

The FLAIR system [2] employed for NER is shown in the Figure 1 is composed by three main components; first a character based Language model (LM) which generates very powerful character based contextual word representations, that are afterwards concatenated with static embeddings. On the top of this LM layer, a BiLSTM layer captures the sequential dependencies among the words of the input sequence. And finally, a conditional random fields (CRF) layer to handle the tagging inference.

The LM in this case concatenates: Static Embeddings and Contextual Flair Embeddings. Contextual FLAIR Embeddings are formed from the character based partial calculations, the BiLSTM strategy is used to take context into account. Those calculations are performed as seen in the bottom of the Figure 1, the results are concatenated with Static Embeddings.
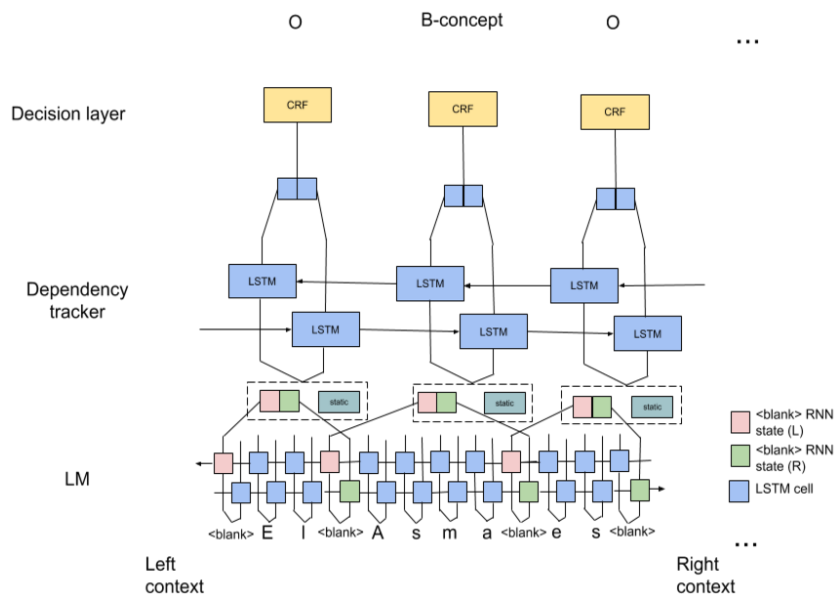
Figure 1: General Architecture of the NER system.

The input provided by the organizer as BRAT standoff format, was tokenized using NLTK word_tokenize general purpose function and afterwards converted to the Inside Outside Beginning format (IOB). This format does not capture overlapped and disjoint entities. The development set was divided so, one part was used for development and other for test. The output of the system was converted to the required BRAT standoff format into the corresponding (.ann) files, this consists on an entity, offsets and the matched text.

3.2. Learning Setup

The proposed architecture was submitted in the current approach, the language model was composed by Contextual and Static Word Embeddings, in between the LM and the Dependency layer was proposed a dropout layer, finally the Prediction layer was connected after it. The architecture used for training can be seen on the Table 1. The training hyperparemeters can summarized as follows: learning rate 0.1, batch size 16 and patience 3 for early stop, it takes into account over-fitting in development file. Maximum of hundred epochs of training were performed, and stopped at 81 epoch. The process was computed on CPU AMD Ryzen 7 1700 Eight-Core Processor, and took 45 minutes to end.

In the current experiment we used pre-trained FastText Static Embeddings (es-crawl) [21] trained over Web crawls (general-domain) and Contextual Flair Embeddings (es-forward + es-backward) [2] trained with Wikipedia (general-domain). All embedding layer were calculated keeping the default parameters. We did not use the additional Medline sentences to train the LM. Therefore no in-domain fine-tunning was pursued.

Layer	Dropout	Specification
LM-Forward	0.5	Embedding (275, 100)
	-	LSTM (100, 2048)
	-	Linear (in = 2048, out = 275, bias=True)
LM-Backward	0.5	Embedding (275, 100)
	-	LSTM (100, 2048)
	-	Linear (in = 2048, out = 275, bias=True)
	-	Embedding (275, 100)
Dependency Tracker	0.5+0.05(word)	Linear (in = 4396, out = 4396, bias=True)
	-	BiLSTM (in= 4396, out = 600)
Decision layer	-	Linear (in = 600, out = 11, bias=True, crf = true)

Table 1
NER hyperparameter setting.

4. Relation Extraction system

On this section we describe our relation extraction (RE) component. In total we have built three RE systems: XLMem, XLMem* and XLMem*+MTB. All the models are based on the same XLM with entity-markers (XLMem) architecture, but they differ on training strategies and data. We first describe the base architecture of XMLem models. In the following sections, we discuss different training strategies and the hyperparameter values used in training.

4.1. XLMem Architecture

The basic architecture of our system is the relation encoder. The encoder consist on a transformer [15] based pre-trained language model with a relation extraction head on top. A particularity of this relation encoder is the need of Entity Markers [4] as additional tokens on the input sentence. This special tokens delimits the boundaries of each entity on the input sentence as shown in the Figure 2. The entity aware input is then fed to the pre-trained language model. The relation extraction head concatenates the representations of the markers that indicate the starting position of the entities and combine them with a linear layer encoding the final relation representation.

Formally, given a relation statement $r = (x, e_1, e_2)$ formed by a sequence of tokens $x = [x_0, x_1, ..., x_n]$ and two entities e_1 and e_2, we first corrupt our input sentence by adding the entity markers ([E1S] and [E1E] defines where the first entity **starts** and **ends**)

$$\tilde{x} = [x_0, ..., [E1S], e_1, [E1E], ..., [E2S], e_2, [E2E], ..., x_n]$$

then we obtain the hidden representations $h = Transformer(\tilde{x})$ and finally we define our relation encoder f_θ as follows:
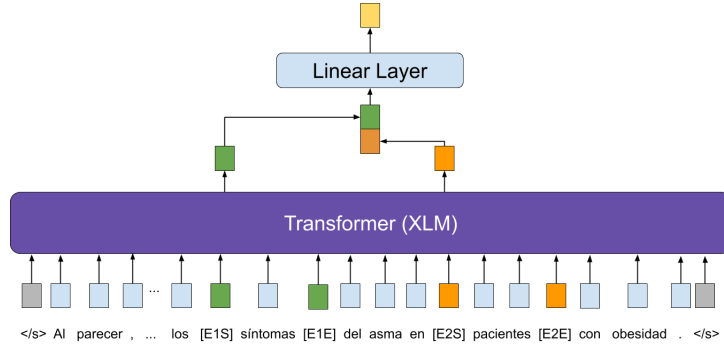
Figure 2: The XLMem relation encoder architecture based on Entity Marker strategy for relation representation.

$$f_\theta = W_{re}[h_{E1S}; h_{E2S}] + b_{re} \qquad (1)$$

where $W_{re} \in \mathbb{R}^{2H \times H}$ and $b_{re} \in \mathbb{R}^H$ being H the hidden representation size. Finally, classification is performed by stacking a linear layer on top of the f_θ encoder with a softmax activation function:

$$output(r) = softmax(W_{clf}f_\theta(r) + b_{clf}) \qquad (2)$$

where $W_{clf} \in \mathbb{R}^{K \times H}$ and $b_{clf} \in \mathbb{R}^K$ being H the hidden representation size and K the number of relations.

Using XLM as pre-trained language model gives the opportunity to learn a cross-lingual relation-encoder, which seems to be a good choice for this setting. Concretely, we use the *xlm-mlm-17-1280* checkpoint provided by the Hugging Face team [22]. This particular checkpoint has been trained on the Masked Language Model (MLM) strategy with 17 languages including Spanish, which is our target language for the task.

4.2. Matching the Blanks

Matching the Blanks [4] (MTB) can be seen as novel alternative to the well-known Distant Supervision [18]. The approach is based on the hypothesis that if two entities are related, sentences that contains those two entities are more likely to express same relation. The Figure 3 shows three different sentences from our MTB corpus. While the first two sentences encode the same relation for *paciente* and *síntomas*, the third sentence express the relation for *paciente* and *tiempo*.

Training dataset is generated as follows. We generate positive pairs of sentence (i.e examples 1 and 2 in Figure 3) if they share blanked entities. We generate strong negative pairs if they share one entity (i.e examples 1 and 3), and weak negatives if no entity is shared. Once we have generated those examples, we train a model that learns whether a pair of sentences encodes the same relation or not, and we transfer learned parameters to the actual relation extraction task.Note that [blank] are introduced to avoid simply relearning a linking entities to Knowledge Base (KB) used to generate the MTB corpus.

(1) Se observó actividad de CK en [BLANK] con dengue con presencia de [BLANK] como vómito, hematemesis y dolor abdominal.

(2) Al parecer, existen mecanismos comunes a ambas patologías que pueden influir en la exacerbación de los [BLANK] del asma en [BLANK] con obesidad.

(3) El [BLANK] promedio para el inicio de ENT fue de 30 (23,5) horas, y el 88,7% de los [BLANK] alcanzaron el objetivo nutricional en 48 horas.

Figure 3: Three different entries on the MTB dataset. The first two share the same entities *paciente* and *síntomas* and the third that contains the entities *paciente* and *tiempo* shares only one of them.

Hyperparameter	MTB pre-training	fine-tuning
Learning-rate	$1e^{-4}$	$3e^{-4}$
Optimizer	SGD	SGD
Batch size	8	16
Gradient accumulation steps	8	4
Floating point precision	FP16 and FP32	FP16 and FP32
Early stopping (patience 3)	✓	✓
Blanks masking probability	0.7	-

Table 2
Hyperparameter settings of the Relation Extraction system training process.

To build the MTB corpus we have use spanish Medline abstracts. We have processed them with Freeling [23] to extract medical entities. In total we got $7,543$ medical entities that forms $278,956$ entity-pairs. With those entity-pairs we have generated $691,392$ positive instances and $833,332$ negative instances. We have split our data into 80% for training and 20% for development. Finally, for technical reasons we discard those instances that contains contexts larger than 128 tokens.

4.3. Learning Setup

In this section we report the set of hyperparameters that we have used at time of fine-tuning the models (Figure 2). For our case the hyperparameters that better fits to the development set were the same for the three tested approaches. Also we report the hyperparameters used during MTB pre-training.

The reported configurations are used on a single NVIDIA Titan V GPU with 12Gb of RAM. The fine-tune process takes less than 10 hours and in the case of MTB pre-training we have manually stopped for reasons of time and deadlines.

5. Results

Table 3 shows the results for the main and alternative domain tasks. Our official run combines the NER system with the XLMem* RE system. In this case XLMem* make use of the additional

Model	Main eval.			Alternative		
	Prec.	Rec.	F1	Prec.	Rec.	F1
Vicomtech	**0.679**	**0.652**	**0.665**	0.594	0.535	0.563
Talp-UPC	0.626	0.626	0.626	0.604	**0.563**	**0.583**
UH-MAJA-KD	0.634	0.615	0.625	**0.608**	0.498	0.547
IXA-NER-RE	0.536	0.580	0.557	0.563	0.416	0.478

Table 3
Official results of the best systems in the main and alternative domain tasks.

Model	Dev			Test		
	Prec.	Rec.	F1	Prec.	Rec.	F1
SINAI	-	-	-	**0.845**	0.807	**0.825**
Vicomtech	-	-	-	0.822	0.820	0.821
Talp-UPC	-	-	-	0.807	**0.825**	0.816
UH-MAJA-KD	-	-	-	0.820	0.808	0.814
UH-MatCom	-	-	-	0.795	**0.825**	0.767
baseline	-	-	-	0.542	0.504	0.586
(Ours) BiLSTM + CRF	0.742	0.696	0.718	0.692	0.727	0.660

Table 4
Results for Named Entity Recognition task. The BiLSTM + CRF system is the one that has been send to the competition.

3,000 automatically annotated sentences from Medline that were provided for further training. Overall results show that although our system is competitive (4th overall rank) still has large room of improvement in the main task, as well as in the alternative task. We would like to note that further in-house evaluation showed that our best system combination would be using XLMem without using extra automatic annotated data.

5.1. Entity Extraction Task (A)

In the following Table 4 could be seen the Test results of the NER task, the best results for each metric are highlighted with bold characters, we also provide our results over Dev set, and over the official Test set.

Although far from the result obtained by the first classified, the system presented overcomes the baseline with no fine-tunning and using general-domain static embeddings and pre-trained Language Model. A preliminary error analysis has led us to conclude that contrary to what we initially thought the domain might be relevant in this NER task. Although 3 of the four types of entities (*actions*, *references* and *predicates*) are not specially medical domain entity types, the fact that predicting *references* and *predicates* is strongly conditioned to having previously correctly predicted their antecedent *concept*, and the latter is most of the times domain specific

Model	Train			Dev			Test		
	Prec.	Rec.	F1	Prec.	Rec.	F1	Prec.	Rec.	F1
Vicomtech	-	-	-	-	-	-	0.672	0.515	0.583
UH-MAJA-KD	-	-	-	-	-	-	0.629	0.571	0.599
(Ours) XLMem	**0.861**	**0.849**	**0.855**	0.708	0.642	0.674	0.690	**0.625**	**0.656**
(Ours) XLMem*	0.767	0.795	0.781	0.707	**0.672**	**0.689**	0.649	0.619	0.633
(Ours) XLMem*+MTB	0.788	0.709	0.746	**0.755**	0.616	0.678	**0.707**	0.584	0.640

Table 5
Results obtained by the different systems in the Relation Extraction task. Best results on each metric are marked in bold and * indicates the use of extra automatic annotated data for training. The XLMem* system is the one that has been send to the competition.

might have an impact.

5.2. Relation Extraction Task (B)

On this part we discuss the the results obtained by our RE systems during the development and testing. We compare our three different systems with the rest top competitors and we evaluate more exhaustively our own systems by comparing the precision-recall curves and analyzing the confusion in the prediction.

Table 5 shows a comparative results (precision, recall, and F1 score) between our systems and the rest top competitors. As only the Test set was reported by the organizers, we just compare our system with the rest on that specific partition. Results on the development show the following: 1) the additional automatic annotated data have positive effect on the model regularization, and 2) MTB pre-training gives a boost on the precision by loosing on the recall. The best model according to the development set is the **XLMem***, which was part of the official run. On the contrary, Test results show unexpected behaviour. We hypothesize this is due to the differences on the relation-type distribution of development and test partitions (Figure 4). Nevertheless, each of the proposed relation extractors outperforms the rest of the systems by a large margin.

Figure 5 reports precision-recall curves on relation categories of the three RE models. The curves show that the XLMem and XLMem* systems performs similar as their *micro-averaged* curves are very close, but not for the XLMem*+MTB, which the curve performs under the rest. On the other hand, curves on relation categories show that the XLMem* system performs better in the *action-roles* relations, and the XLMem perform better in the *general* domain relations. The differences on development and test can also be explained by the distribution shown in Figure 4. Finally, analysis of the output reveals that the confusion is located actually between the negative class (no-relation) and the rest of positive relations (i.e. false negatives), and not between the positive relation types.
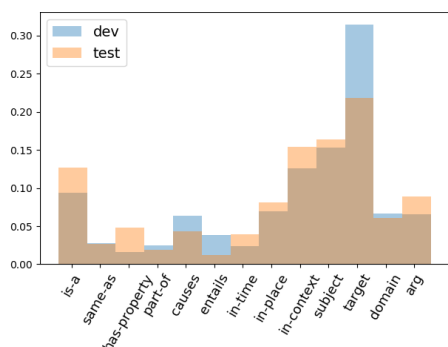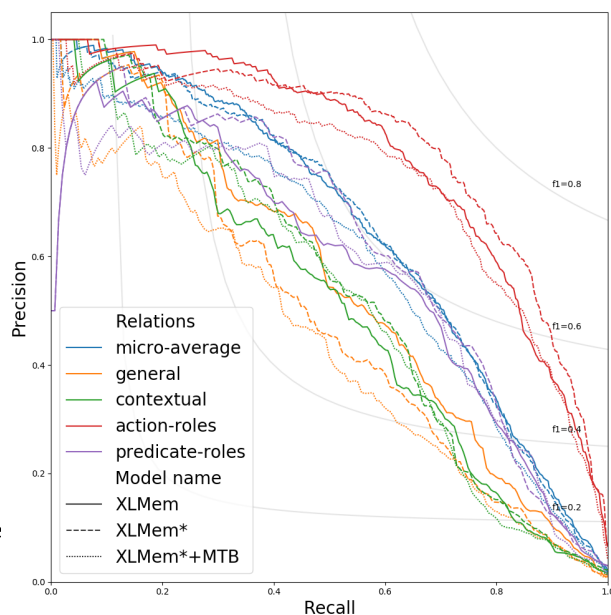
Figure 4: Relation distribution of developme

Figure 5: Precision/Recall curves of the different systems.

6. Conclusions

The purpose of this work was to evaluate the feasibility of different approaches to medical entity recognition and relation extraction for Spanish. Entity recognition was approached with a character based sequence labeler, and for the relation extraction we used a fine-tuned large multilingual pre-trained language model. Proposed system shows promising results. We ranked 4th overall, and obtain the best results for the relation extraction task. In the future, we plan to improve the entity recognition part by means of using a domain specific LM, and further investigate the use of Matching the Blanks method as a data-augmentation technique.

References

[1] A. Piad-Morffis, Y. Gutiérrez, H. Cañizares-Diaz, S. Estevez-Velarde, Y. Almeida-Cruz, R. Muñoz, A. Montoyo, Overview of the eHealth Knowledge Discovery Challenge at IberLEF 2020, in: Proceedings of the Iberian Languages Evaluation Forum co-located with 36th Conference of the Spanish Society for Natural Language Processing, IberLEF@SEPLN 2020, Spain, September, 2020., 2020.

[2] A. Akbik, D. Blythe, R. Vollgraf, Contextual string embeddings for sequence labeling, in: Proceedings of the 27th International Conference on Computational Linguistics, Association for Computational Linguistics, Santa Fe, New Mexico, USA, 2018, pp. 1638–1649. URL: https://www.aclweb.org/anthology/C18-1139.

[3] G. Lample, A. Conneau, Cross-lingual language model pretraining, Advances in Neural Information Processing Systems (NeurIPS) (2019).

[4] L. Baldini Soares, N. FitzGerald, J. Ling, T. Kwiatkowski, Matching the blanks: Distributional similarity for relation learning, in: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, Florence, Italy, 2019, pp. 2895–2905. URL: https://www.aclweb.org/anthology/P19-1279. doi:10.18653/v1/P19-1279.

[5] G. Zhou, J. Zhang, J. Su, D. Shen, C. Tan, Recognizing names in biomedical texts: a machine learning approach, Bioinformatics 20 (2004) 1178–1190.

[6] F. Soares, M. Villegas, A. Gonzalez-Agirre, M. Krallinger, J. Armengol-Estapé, Medical word embeddings for Spanish: Development and evaluation, in: Proceedings of the 2nd Clinical Natural Language Processing Workshop, Association for Computational Linguistics, Minneapolis, Minnesota, USA, 2019.

[7] P. Stenetorp, H. Soyer, S. Pyysalo, S. Ananiadou, T. Chikayama, Size (and domain) matters: Evaluating semantic word space representations for biomedical text, in: Proceedings of the 5th International Symposium on Semantic Mining in Biomedicine, Zürich, Switzerland, 2012.

[8] M. E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, L. Zettlemoyer, Deep contextualized word representations, in: Proc. of NAACL, 2018.

[9] J. Howard, S. Ruder, Universal language model fine-tuning for text classification, 2018. arXiv:1801.06146.

[10] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, arXiv preprint arXiv:1810.04805 (2018).

[11] L. Akhtyamova, P. Martinez, K. Verspoor, J. Cardiff, Testing contextualized word embeddings to improve ner in spanish clinical case narratives, BMC Medical Informatics and Decision Making (2020) preprint. doi:10.21203/rs.2.22697/v1.

[12] J. Lee, W. Yoon, S. Kim, D. Kim, S. Kim, C. H. So, J. Kang, BioBERT: a pre-trained biomedical language representation model for biomedical text mining, Bioinformatics (2019). URL: https://doi.org/10.1093/bioinformatics/btz682. doi:10.1093/bioinformatics/btz682.

[13] Y. Si, J. Wang, H. Xu, K. Roberts, Enhancing clinical concept extraction with contextual embeddings, Journal of the American Medical Informatics Association 26 (2019) 1297–1304. URL: http://dx.doi.org/10.1093/jamia/ocz096. doi:10.1093/jamia/ocz096.

[14] G. Sheikhshabbafghi, I. Birol, A. Sarkar, In-domain context-aware token embeddings improve biomedical named entity recognition, in: Proceedings of the Ninth International Workshop on Health Text Mining and Information Analysis, Association for Computational Linguistics, Brussels, Belgium, 2018, pp. 160–164. URL: https://www.aclweb.org/anthology/W18-5618. doi:10.18653/v1/W18-5618.

[15] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, in: Advances in neural information processing systems, 2017, pp. 5998–6008.

[16] M. E. Peters, M. Neumann, R. L. Logan, R. Schwartz, V. Joshi, S. Singh, N. A. Smith, Knowledge enhanced contextual word representations, in: EMNLP, 2019.

[17] M. Joshi, D. Chen, Y. Liu, D. S. Weld, L. Zettlemoyer, O. Levy, SpanBERT: Improving pre-training by representing and predicting spans, arXiv preprint arXiv:1907.10529 (2019).

[18] M. Mintz, S. Bills, R. Snow, D. Jurafsky, Distant supervision for relation extraction without labeled data, in: Proceedings of the Joint Conference of the 47th Annual Meeting of the

ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP, Association for Computational Linguistics, Suntec, Singapore, 2009, pp. 1003–1011. URL: https://www.aclweb.org/anthology/P09-1113.

[19] O. Sainz, O. Lopez de Lacalle, I. Aldabe, M. Maritxalar, Domain adapted distant supervision for pedagogically motivated relation extraction, in: Proceedings of The 12th Language Resources and Evaluation Conference, European Language Resources Association, Marseille, France, 2020, pp. 2213–2222. URL: https://www.aclweb.org/anthology/2020.lrec-1.270.

[20] R. Hoffmann, C. Zhang, X. Ling, L. Zettlemoyer, D. S. Weld, Knowledge-based weak supervision for information extraction of overlapping relations, in: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Association for Computational Linguistics, Portland, Oregon, USA, 2011, pp. 541–550. URL: https://www.aclweb.org/anthology/P11-1055.

[21] E. Grave, P. Bojanowski, P. Gupta, A. Joulin, T. Mikolov, Learning word vectors for 157 languages, in: Proceedings of the International Conference on Language Resources and Evaluation (LREC 2018), 2018.

[22] T. Wolf, L. Debut, V. Sanh, J. Chaumond, C. Delangue, A. Moi, P. Cistac, T. Rault, R. Louf, M. Funtowicz, J. Brew, Huggingface's transformers: State-of-the-art natural language processing, ArXiv abs/1910.03771 (2019).

[23] L. Padró, E. Stanilovsky, Freeling 3.0: Towards wider multilinguality, in: Proceedings of the Language Resources and Evaluation Conference (LREC 2012), ELRA, Istanbul, Turkey, 2012.