

# Bete: A Brazilian Portuguese Dataset for Named Entity Recognition and Relation Extraction in the Diabetes Healthcare Domain

Lucas Pavanelli<sup>1</sup>[0000-0003-2228-7965], Yohan Bonescki Gumiel<sup>2</sup>[0000-0001-8239-2930], Thiago Ferreira<sup>1</sup>[0000-0003-0200-3646], Adriana Pagano<sup>3</sup>[0000-0002-3150-3503], and Eduardo Laber<sup>4</sup>[0000-0002-9025-8333]

<sup>1</sup> aiXplain Inc.

{lucas.pavanelli,thiago}@aixplain.com

<sup>2</sup> Universidade Federal de Minas Gerais

apagano@ufmg.br, yohan.gumiel@gmail.com

<sup>3</sup> Pontifícia Universidade Católica do Rio de Janeiro (PUC-RJ)

laber@inf.puc-rio.br

**Abstract.** The biomedical NLP community has seen great advances in dataset development mostly for the English language, which has hindered progress in the field, as other languages are still underrepresented. This study introduces a dataset of Brazilian Portuguese annotated for named entity recognition and relation extraction in the healthcare domain. We compiled and annotated a corpus of health professionals’ responses to frequently asked questions in online healthcare forums on diabetes. We measured inter-annotator agreement and conducted initial experiments using up-to-date methods to recognize entities and extract relations, such as BERT-based ones. Data, models, and results are publicly available at <https://github.com/pavalucas/Bete>.

**Keywords:** Dataset · Named Entity Recognition · Relation Extraction.

## 1 Introduction

Named entity recognition (NER) identifies named entities in texts; these entities can be proper names, locations, and organizations. Relation extraction (RE) consists in predicting whether and what predefined relation exists between two mentioned entities or if they are unrelated. In the healthcare domain, named entities include mentions relevant to the clinical context, such as treatment plans, drugs, and symptoms while relations can be, for instance, if a certain drug treats a symptom. Both tasks are prevalent in machine learning methods [2] and are considered the primary step for several other Natural Language Processing (NLP) tasks, such as question answering [20], document classification [6], and search engines [12].

Consultation time is not always sufficient for health professionals to answer all the patient and their families questions, and patients may not have easy access

to primary care centers [8]. Hence, patients often engage in dedicated public forums to search for answers to their queries. Among chronic diseases, diabetes stands out as a condition much in need of attention because it is an increasingly prevalent and severe long-term problem quickly growing worldwide, particularly in developing countries, as is the case of Brazil. In 2019, it was estimated that almost half a billion of the world’s population (9.3% of the adults between 20-79 years) had diabetes [18]. Further, one in every two (50.1%) persons with diabetes is unaware of or has not been diagnosed with this condition. Hence, question answering (QA) systems could help relieve some of the burdens of health care. In QA, an answer to a question is found by querying a corpus of text documents. A QA system incorporates NER and RE components so that entities and their relations can be detected and answers can be more efficiently found.

In this study, the main goal was to annotate a corpus of texts in order to develop a framework to automatically identify healthcare-related entities and relations. To that end, medical and nutrition science students produced texts as answers to queries posted by users in diabetes-related public forums. Our work is thus relevant for studies in the healthcare domain targeting specifically the general public and focusing on Brazilian Portuguese. The obtained dataset is expected to be used to build BeteQA, a community question answering system that provides fast and precise answers to questions about Diabetes Mellitus posed by the lay public [5].

To the best of our knowledge, this is the first study focusing on NER and relation extraction drawing on a novel corpus of real-life questions answered by professional healthcare specialists in Brazilian Portuguese.

## 1.1 Related Work

Among named entity recognition methods, contextual word representations are employed due to their capacity for modeling complex word usage characteristics and variations across linguistic contexts [17]. A method worth highlighting is BERT [7], a masked language model that benefits from transformers. These pre-trained language models are becoming the new NER paradigm because of their contextualized embeddings. Furthermore, they are fine-tuned for several NLP tasks by adding an additional output layer [13].

As far as relation extraction methods, BERT-based models are also dominant. Soares et al. [21] proposed a model that learns relation representations directly from the text using BERT-based models.

Considering the available BERT models for Brazilian Portuguese, there is a multilingual version [7]; a Brazilian Portuguese version focused on the general domain called BERTimbau [22]; and a Brazilian Portuguese version focused on the clinical domain named BioBERTpt [19].

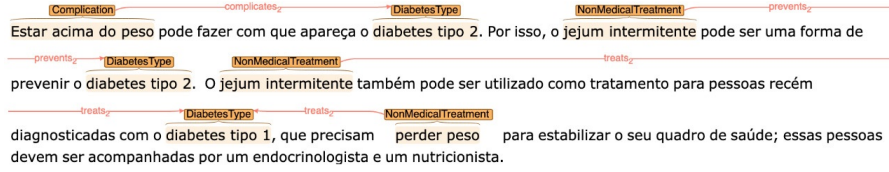


Fig. 1. Annotation example from our corpus.

## 2 Methodology

### 2.1 Corpus

We searched for questions posed by users regarding health issues in several online forums. In such forums, users answer each other’s questions drawing on their beliefs and understanding of the issues, with no help or supervision of any healthcare professional. To avoid this problem, we designed a study in which answers were produced by medical and nutrition science students relying on their domain knowledge. Answers were curated by expert professionals in our project who supervised students. This way, we ensured that our set of QA is reliable and can be used in a prospective QA system to be queried through a conversational agent. Our corpus comprises two sets of documents made up of diabetes-related general questions with their respective answers: a first set contains 304 real online forum questions answered by medical students under the supervision of medical professionals and a second one contains 201 real online forum questions answered by nutrition science students supervised by professionals. This way, we can have a QA system to provide accurate answers about health and nutrition issues, authored by professionals in both fields.

**Annotation Setup** As an annotation tool, we used Webanno [4], an open-source and intuitive software/platform. After comparing several available entity tagging tools, we found Webanno to be the easiest and most efficient tool for our purposes. The system was set up on a web server, text data was uploaded, and entity/relation types were defined within the system.

**Annotation Guidelines** The annotation guidelines were created in an iterative process. A first draft was created, containing general guidelines as well as specific examples of types of entities and relations. Specialists were consulted regarding annotators’ queries and their answer was used to update the guidelines, then they were tested again. Besides, during the annotation process, whenever one of the annotators ran into a dubious case, this was added to the guidelines. Figure 1 shows an example of a text annotated in Webanno following our guidelines<sup>4</sup>. The annotation guidelines are publicly available for download as part of the dataset.

<sup>4</sup> Translation into English: “Being overweight can lead to type 2 diabetes. Therefore, intermittent fasting may be a way to prevent type 2 diabetes. Intermittent fasting

**Annotation Process** We recruited undergraduate students pursuing their BA degree to complete the annotation task. The students are part of Empoder@, a multidisciplinary project engaging health sciences, statistics, computer science, and applied linguistics students. The project aims to empower researchers, professionals, and users of health services.

Annotators took part in a training session and were requested to read the guidelines and resort to the project coordinator whenever they encountered problems during annotation. Two students annotated each document, and a third one performed the adjudication so that a gold-standard was obtained upon completion. Annotators were presented with each text on a simple interface (see Figure 1) and using a mouse or a track pad they selected entities and dragged relationships between them.

## 2.2 Entity and Relation Extraction

We considered relation extraction as a multi-label classification problem in which, given a pair of entities, a label is assigned out of the relation types available. Also, we decoupled relation extraction from entity recognition, so we performed RE on the gold entities.

The set of entity and relation types devised for our annotation was built drawing on an ontology proposed in [1]. The ontology labels diabetes mentions as “DiabetesType”, and diabetes-related diseases are classed with the “Complication” entity type. Diabetes-related temporal expressions, clinical tests, and treatments are also addressed in the ontology. Table 1 lists the 14 entity types, providing a brief explanation of each label with some examples. The 5 relations types are verbalized by the following verbs: “causes”, “diagnoses”, “has”, “prevents”, and “treats”.

**Reliability** To measure dataset reliability, we computed inter-annotator agreement (IAA) considering exact matches. Following the work of [3], we computed pairwise F1 score and Cohen’s Kappa. The former is more reliable according to various studies [11, 9]. Because of the vast amount of unannotated tokens (labeled “O”), we calculated the scores without the O label, for both annotated entities and relations. Table 2 shows the obtained agreement. Annotated entities can be said to be fully reliable, achieving an IAA of 0.93. As regards relations, moderate agreement (0.58) was found.

## 3 Dataset information

Table 3 shows overall statistics of the whole dataset. Table 4 shows the number of annotations per entity type, while table 5 covers the annotated relations.

---

can also be used as a treatment for people newly diagnosed with type 1 diabetes who need to lose weight to achieve a more stable health condition; these people should be advised and monitored by an endocrinologist and a nutritionist.”

**Table 1.** Entities description and examples.

Category	Description	Examples of annotated entities
Diabetes Type	subclass of diabetes	type 2 – type 1
Complication	diseases and health conditions causing or caused by diabetes	being overweight – wounds – depression - neuropathy
Symptom	physical or mental condition experienced by the patient regarded as indicating diabetes	low blood sugar
Glucose Value	measurement of blood sugar level	250 – 100 - 80
Insulin	insulin type	NPH – Aspart
Medication	prescribed drugs or medicine	Metformin - Tetracaine hydrochloride
Non Medical Treatment	healthcare activities or behavior other than prescribed medication	intermittent fasting – physical exercise
Food	source of nutritional support for organisms	peanut butter, candies, bread
Dose	amount, quantity or size	150ml – 200g – 1 glass
Test	medical exams	blood test - glycosylated hemoglobin test
Date	calendar dates	17/01/2021
Time	point in time	at night – at bedtime – at midday
Duration	length of time for occurrence	half an hour – twenty minutes
Set	frequency of occurrence	twice a week – every day

**Table 2.** Inter-annotator agreement for entity recognition and relation extraction.

Annotation type	F1 score	Cohen’s Kappa
Entity	0.93	0.91
Relation	0.58	0.32

**Table 3.** Dataset information.

Documents	Sentences	Tokens	Entities	Relations
505	2340	55530	2396	1223

**Table 4.** Entities: Number of occurrences and percentage per entity type sorted in decreasing order.

<b>Entity</b>	<b>Count</b>	<b>%</b>
Food	631	26.35
Complication	410	17.12
NonMedicalTreatment	405	16.91
Symptom	309	12.90
GlucoseValue	308	12.85
Time	71	2.96
Test	67	2.80
Medication	52	2.17
DiabetesType	48	2.00
Set	29	1.21
Insulin	25	1.04
Dose	23	0.96
Duration	18	0.75

**Table 5.** Relations: Number of occurrences and percentage per relation type sorted in decreasing order.

<b>Relation</b>	<b>Count</b>	<b>%</b>
has	833	68.11
treats	202	16.52
causes	79	7.03
diagnoses	52	4.25
prevents	50	4.09

## 4 Experiments Setup

We conducted initial experiments using methods to recognize entities and extract relations. Since the dataset is in Brazilian Portuguese, we chose deep learning models trained on multilingual and Brazilian Portuguese data. These models are multilingual BERT (mBERT), BERTimbau, and the three different versions of BioBERTpt: BioBERTpt-bio, trained on Portuguese biomedical texts, BioBERTpt-clin, trained on clinical narratives from electronic health records from Brazilian Hospitals, and BioBERTpt-all, trained in both biomedical texts and clinical narratives.

Regarding the training setup, we randomly divided the 505 documents into train/dev/test using the split 0.8/0.1/0.1, respectively, tuning the hyperparameters on the development set and reporting the results on the test set. To run the experiments, we used one NVIDIA GeForce RTX 3090 GPU.

As for models, we trained a baseline Conditional Random Field (CRF) for the NER task. We used a Portuguese model from the Python library called Spacy [10] to extract a set of features. Table 6 shows the used features.

Feature
Part of speech tagging, extracted using Spacy [10].
If the word is in uppercase.
If the word is a digit.
If the word is a title, i.e. start with an uppercase letter and the rest is in lowercase.
The previous word and the above features.
The next word and the above features.

**Table 6.** Used CRF features.

For BERT models for the NER task, we used the Adam [14] optimizer with a learning rate of 1e-5 and a maximum length of 512. Moreover, we trained for 50 epochs with early stopping of 15 epochs and a batch size of 64.

Regarding the relation extraction task, we experimented with a baseline Support Vector Machine (SVM) model, using the one-vs-the-rest (OvR) multiclass strategy from scikit-learn [16] Python library.

Considering BERT for relation extraction, we used 7e-5 as the learning rate with Adam optimizer and max length of 512, training for 11 epochs, and a batch size of 32.

We considered the following metrics for evaluation: precision, recall, and F1 score. We reported the weighted average F1 score, averaging the support weighted mean per label. All metrics are considering exact matches. In addition, we computed results considering all classes and, for the best-performing model, we reported metrics for each label.

## 5 Results

Table 7 shows the results for NER models. We trained baseline CRF and BERT-based models. The best one concerning F1 score is the BioBERTpt-clin model, outperforming BioBERTpt-all by 0.9 points. Also, BERTimbau did not perform well with an F1 score 5.3 points lower than the second-lowest.

To evaluate the relation extraction task, we experimented using the two models: SVM, as the baseline, and BERT for relation extraction (BERT-RE) from [21], using mBERT as the BERT encoder.

Table 8 shows the result for both relation extraction models. We can observe that, although SVM has higher precision than BERT-RE, the latter performs better overall with 17 points of F1 score difference.

We also provide an in-depth analysis of the best-performing model in NER and RE. Table 9 shows detailed results for the entity recognition BioBERTpt-clin model. The method produced good scores (>80%) for the three entities with the largest number of examples: Food, Complication, and Symptom. However, for some entities with few examples, the model was not able to perform good predictions (<60%): Test, Time, and Set.

Table 10 describes an in-depth analysis of BERT-RE performance for each relation type. Similarly to NER results, the model performs well for the relation with most examples (has) and falls short in relations with few examples: causes and prevents.

**Table 7.** Experiments for entity recognition models. The best scores are highlighted in bold.

Model	Precision	Recall	F1
CRF	<b>80.3</b>	72.9	76.1
BioBERTpt-bio	73.1	80.5	76.6
BioBERTpt-clin	77.5	81.8	<b>79.4</b>
BioBERTpt-all	74.5	<b>83.2</b>	78.5
BERTimbau	72.2	70.2	70.8
mBERT	74.3	81.6	77.6

**Table 8.** Experiments for relation extraction models. The best scores are highlighted in bold.

Model	Precision	Recall	F1
SVM	<b>80.0</b>	46.2	58.1
BERT-RE (mBERT)	73.6	<b>77.8</b>	<b>75.1</b>



**Table 9.** Detailed results for the best entity recognition model (BioBERTpt-clin).

Entities	Precision	Recall	F1	N° of examples
Complication	84.8	91.8	88.2	61
DiabetesType	100.0	100.0	100.0	2
Dose	50.0	100.0	66.7	2
Duration	100.0	100.0	100.0	1
Food	79.1	81.5	80.3	135
GlucoseValue	60.4	78.4	68.2	37
Insulin	100.0	100.0	100.0	4
Medication	68.8	91.67	78.6	12
NonMedicalTreatment	72.7	64.9	68.6	37
Set	66.7	50.0	57.1	4
Symptom	90.0	94.7	92.3	57
Test	50.0	33.3	40.0	9
Time	44.4	50.0	47.1	8
<b>Weighted Average</b>	<b>77.5</b>	<b>81.8</b>	<b>79.4</b>	<b>369</b>

**Table 10.** Detailed results for the best relation extraction model (BERT-RE).

Entities	Precision	Recall	F1	N° of examples
causes	33.33	33.33	33.33	6
prevents	66.67	50.00	57.14	4
treats	52.94	64.29	58.06	14
has	77.89	86.05	81.77	86
diagnoses	100.00	57.14	72.73	7
<b>Weighted Average</b>	<b>73.56</b>	<b>77.78</b>	<b>75.06</b>	<b>117</b>

## 6 Discussion

In this study, we introduced a novel dataset and models for NER and RE tasks, gathering answers written by health professionals about Diabetes Mellitus to online forum users.

**Corpus** We found that the entities Food and Complication were the most annotated in the corpus. Also, nutrition is a prevalent topic in online forums, as people with diabetes frequently inquire whether they can or cannot consume specific foods, such as chocolate, or alcoholic drinks. Similarly, people are generally keen on finding further information on diabetes-related complications because they feel specific symptoms or want to know about frequent diabetes-related complications.

Regarding the main difficulties found in the process of annotating relations, despite annotators’ prior training and availability of guidelines, some entities were annotated as having a different extent in terms of words making up those entities. Thus, “type 2 Diabetes” and “Diabetes” are two different annotations

that refer to the same entity type, “DiabetesType”. This ends up impacting the relation since a relation is defined by a pair of entities and a relation type, which can justify our lower agreement for relation annotation. So if two annotators annotated the same relation type holding between entities spanning different extents in words, the resulting relations did not match. Further, as evidenced in [15], temporal relation extraction, which is a specific type of relation extraction, usually has lower agreement than span annotation.

**Models** For the NER task, among the BERT models, we found that BioBERTpt-clin, which leveraged clinical data, had superior performance compared to the other models. Both BioBERTpt-clin and BioBERTpt-all were trained on clinical data and were initialized with the weights from Multilingual BERT, so as expected, these models achieved similar results. It is also worth noting that, although trained on Brazilian Portuguese data, BERTimbau did not perform well. An explanation is that BERTimbau was trained on brWAC [23], a large corpus extracted from the Web, that differs in vocabulary and context from Bete medical data.

Considering relation extraction, the BERT-based model outperforms the baseline model, showing the dominance of context-aware embeddings compared to kernel-based methods that were dominant in the past.

## 7 Conclusion

To the best of our knowledge, this is the first study yielding an annotated corpus of NER in Brazilian Portuguese made up of diabetes-related answers authored by domain specialists in response to questions posed by lay users. Moreover, it contributes to the research field by introducing resources for a sensitive context, such as diabetes, and creating models for a low-resource language, such as Brazilian Portuguese.

The fine-tuning of models leveraging clinical data was found to improve the results. Hence, the vocabulary and the context from the clinical context boosted the model’s ability to predict entities.

We plan to expand our corpus in future work, especially for the categories of entities and relations that have a small number of instances.

## Limitations

Among the limitations of our study is the size of our dataset, which is due to the fact that it was obtained through manual annotation thereby demanding human effort and more time to accomplish this task. Another limitation is that some entity types have few instances, making our dataset slightly imbalanced. For instance, only 1/3 of our entity types have percentages of occurrence higher than 3%. Hence, adding more training examples and making the dataset less imbalanced will certainly enhance our results.

Our dataset targets a particular domain - diabetes; future experiments targeting other chronic diseases, such as cardiovascular disease, will boost our potential. Additionally, in the aftermath of COVID-19, there is growing uncertainty about procedures, symptoms, and treatments, with several questions being asked over social media. Hence, addressing COVID-19-related frequently asked questions would be a valuable contribution.

## Ethics Statement

Our study fully complies with ethical standards and did not require any submission to ethical boards, since no data collection with human subjects was carried out. Our dataset was created by our team and contains texts drafted by medical students under the supervision of healthcare professionals, all of whom are research members in our project.

## References

1. Ben Abacha, A., Zweigenbaum, P.: Means: A medical question-answering system combining nlp techniques and semantic web technologies. *Information Processing Management* **51**(5), 570–594 (2015). <https://doi.org/https://doi.org/10.1016/j.ipm.2015.04.006>, <https://www.sciencedirect.com/science/article/pii/S0306457315000515>
2. Bose, P., Srinivasan, S., Sleeman, W.C., Palta, J., Kapoor, R., Ghosh, P.: A survey on recent named entity recognition and relationship extraction techniques on clinical texts. *Applied Sciences* **11**(18) (2021). <https://doi.org/10.3390/app11188319>, <https://www.mdpi.com/2076-3417/11/18/8319>
3. Brandsen, A., Verberne, S., Wansleeben, M., Lambers, K.: Creating a dataset for named entity recognition in the archaeology domain. In: *Proceedings of the Twelfth Language Resources and Evaluation Conference*. pp. 4573–4577. European Language Resources Association, Marseille, France (May 2020), <https://aclanthology.org/2020.lrec-1.562>
4. Eckart de Castilho, R., Mújdricza-Maydt, É., Yimam, S.M., Hartmann, S., Gurevych, I., Frank, A., Biemann, C.: A web-based tool for the integrated annotation of semantic and syntactic structures. In: *Proceedings of the Workshop on Language Technology Resources and Tools for Digital Humanities (LT4DH)*. pp. 76–84. The COLING 2016 Organizing Committee, Osaka, Japan (Dec 2016), <https://www.aclweb.org/anthology/W16-4011>
5. Castro Ferreira, T., Victor de Pinho Costa, J., Rigotto, I., Portella, V., Frota, G., Luisa A. R. Guimarães, A., Penna, A., Lee, I., A. Soares, T., Rolim, S., Cunha, R., França, C., Santos, A., F. Oliveira, R., Langbehn, A., Hasan Dalip, D., André Gonçalves, M., Bastos Fóscolo, R., Pagano, A.: Evaluating recognizing question entailment methods for a Portuguese community question-answering system about diabetes mellitus. In: *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2021)*. pp. 234–243. INCOMA Ltd., Held Online (Sep 2021), <https://aclanthology.org/2021.ranlp-main.28>
6. Choudhary, A., Arora, A.: Linguistic feature based learning model for fake news detection and classification. *Expert Systems with Applications* **169**, 114171 (2021)

7. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers) pp. 4171–4186 (Jun 2019). <https://doi.org/10.18653/v1/N19-1423>, <https://www.aclweb.org/anthology/N19-1423>
8. Gabarron, E., Bradway, M., Fernandez-Luque, L., Chomutare, T., Hansen, A.H., Wynn, R., Årsand, E.: Social media for health promotion in diabetes: study protocol for a participatory public health intervention design. *BMC Health Services Research* **18**(1), 414 (Jun 2018). <https://doi.org/10.1186/s12913-018-3178-7>
9. Grouin, C., Rosset, S., Zweigenbaum, P., Fort, K., Galibert, O., Quintard, L.: Proposal for an extension of traditional named entities: From guidelines to evaluation, an overview. In: Proceedings of the 5th linguistic annotation workshop. pp. 92–100 (2011)
10. Honnibal, M., Montani, I.: spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing (2017), to appear
11. Hripcsak, G., Rothschild, A.S.: Agreement, the f-measure, and reliability in information retrieval. *Journal of the American medical informatics association* **12**(3), 296–298 (2005)
12. Lahav, D., Falcon, J.S., Kuehl, B., Johnson, S., Parasa, S., Shomron, N., Chau, D.H., Yang, D., Horvitz, E., Weld, D.S., et al.: A search engine for discovery of scientific challenges and directions. *AAAI* (2022)
13. Li, J., Sun, A., Han, J., Li, C.: A survey on deep learning for named entity recognition. *IEEE Transactions on Knowledge and Data Engineering* (2020)
14. Loshchilov, I., Hutter, F.: Decoupled Weight Decay Regularization. In: International Conference on Learning Representations (2019), <https://openreview.net/forum?id=Bkg6RiCqY7>
15. Nikfarjam, A., Emadzadeh, E., Gonzalez, G.: Towards generating a patient’s timeline: Extracting temporal relationships from clinical notes. *Journal of Biomedical Informatics* **46**, S40–S47 (2013). <https://doi.org/https://doi.org/10.1016/j.jbi.2013.11.001>, supplement: 2012 i2b2 NLP Challenge on Temporal Relations in Clinical Data
16. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* **12**, 2825–2830 (2011)
17. Peters, M.E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., Zettlemoyer, L.: Deep contextualized word representations. In: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers). pp. 2227–2237. Association for Computational Linguistics, New Orleans, Louisiana (Jun 2018). <https://doi.org/10.18653/v1/N18-1202>, <https://aclanthology.org/N18-1202>
18. Saeedi, P., Petersohn, I., Salpea, P., Malanda, B., Karuranga, S., Unwin, N., Colagiuri, S., Guariguata, L., Motala, A.A., Ogurtsova, K., Shaw, J.E., Bright, D., Williams, R.: Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the international diabetes federation diabetes atlas, 9th edition. *Diabetes Research and Clinical Practice* **157**, 107843 (Nov 2019). <https://doi.org/10.1016/j.diabres.2019.107843>

19. Schneider, E.T.R., de Souza, J.V.A., Knafo, J., Oliveira, L.E.S.e., Copara, J., Gumiel, Y.B., Oliveira, L.F.A.d., Paraiso, E.C., Teodoro, D., Barra, C.M.C.M.: BioBERTpt - a Portuguese neural language model for clinical named entity recognition. In: Proceedings of the 3rd Clinical Natural Language Processing Workshop. pp. 65–72. Association for Computational Linguistics (Nov 2020). <https://doi.org/10.18653/v1/2020.clinicalnlp-1.7>
20. Sharma, V., Kulkarni, N., Pranavi, S., Bayomi, G., Nyberg, E., Mitamura, T.: BioAMA: Towards an end to end BioMedical question answering system. In: Proceedings of the BioNLP 2018 workshop. pp. 109–117. Association for Computational Linguistics, Melbourne, Australia (Jul 2018). <https://doi.org/10.18653/v1/W18-2312>, <https://aclanthology.org/W18-2312>
21. Soares, L.B., FitzGerald, N., Ling, J., Kwiatkowski, T.: Matching the blanks: Distributional similarity for relation learning. ACL 2019 - 57th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference pp. 2895–2905 (2020). <https://doi.org/10.18653/v1/p19-1279>
22. Souza, F., Nogueira, R., Lotufo, R.: Bertimbau: Pretrained bert models for brazilian portuguese. In: Cerri, R., Prati, R.C. (eds.) Intelligent Systems. pp. 403–417. Springer International Publishing, Cham (2020)
23. Wagner, J., Wilkens, R., Idiart, M., Villavicencio, A.: The brwac corpus: A new open resource for brazilian portuguese (05 2018)