Automatic Section Classification in Spanish Clinical Narratives Using Chunked Named Entity Recognition.

Andrés Carvallo[1,*], Matías Rojas[4], Carlos Muñoz-Castro[1,3,6], Claudio Aracena[2,6], Rodrigo Guerra[2,5], Benjamín Pizarro[3] and Jocelyn Dunstan[3,4,5,6]

[1]National Center for Artificial Intelligence, Santiago, Chile

[2]Faculty of Physical and Mathematical Sciences, Universidad de Chile, Santiago, Chile

[3]Department of Computer Science, Pontificia Universidad Católica de Chile, Santiago, Chile

[4]Institute for Mathematical Computing, Pontificia Universidad Católica de Chile, Santiago, Chile

[5]Center for Mathematical Modeling, Santiago, Chile

[6]Millenium Institute Foundational Research on Data, Santiago, Chile

Abstract

The extraction and classification of important information from Spanish Electronic Clinical Narratives (ECNs) can be challenging due to the complexity of the clinical text and the limited availability of labeled data. In this paper, we introduce a chunked Named Entity Recognition model designed to parse and classify sections of ECNs into predefined categories. The model aims to improve section identification and classification accuracy within ECNs in the context of the IberLEF ClinAIS Task. Our system achieves a promising performance, obtaining a weighted B2 score of .6958, demonstrating its capability to accurately distinguish borders and boundaries between sections. The paper concludes with a comprehensive analysis of the results, discussing potential implications and suggesting directions for further improvements in clinical text analysis.

Keywords
Natural Language Processing, Clinical Narratives, Named Entity Recognition, Section Segmentation

1 Introduction

Electronic Clinical Narratives (ECN) are the predominant way of documenting and evaluating important details concerning a patient's clinical history and progress [1]. These records embody comprehensive details about a patient's prior medical conditions, treatment procedures undertaken, the specific illnesses' progression, and the respective medical interventions prescribed [2]. Beyond their primary function, these narratives serve secondary roles, aiding in tasks such as identifying rare medical events [3], predicting potential readmissions to the hospital [4], and contributing to public health surveillance [5]. Nonetheless, these narratives' dense complexity and expansive scope present a substantial hurdle in their successful segmentation [2]. The manual extraction of critical treatment data and relevant information becomes a demanding, time-intensive task within such extensive textual content [6]. This process involves partitioning the text into semantically distinct labeled segments using a predefined set. The benefits of

✉ afcarvallo@uc.cl (A. Carvallo)

such segmentation are manifold, offering novel insights about entities that vary depending on the section in which they appear [2, 7]. For instance, a past medical condition mentioned in the patient's history could be instrumental in forecasting potential health risks. In contrast, a symptom mentioned in the Evolution section could hint at possible side effects of a treatment regimen [8].

Introduced at IberLEF 2023 [9], the ClinAIS task [10, 11] aims to tackle the challenge of automatically identifying sections within unstructured Spanish clinical documents. The primary focus of this task is the categorization of ECNs into seven predetermined medical sections: *Present Illness*, *Derived from/to*, *Past Medical History*, *Family History*, *Exploration*, *Treatment*, and *Evolution*, predominantly targeting progress notes. The intrinsic complexity of this task arises from the fact that all lexical units within an ECN pertain to a section, with none falling outside the boundary of the seven predefined medical categories. Moreover, this task carries significant value as it helps bridge the linguistic gap in medical informatics resources available for Spanish-speaking nations, facilitating the automation of section identification within ECNs.

This paper presents a novel methodology to enhance the accuracy and efficiency of identifying these predefined medical sections, thereby ensuring more structured and readily accessible data for clinical decision-making. We detail our proposed solution for the ClinAIS shared task: a chunked Named Entity Recognition model. This model identifies the start and end of each of the seven sections in the ECN and then classifies this segment under one of the seven possible labels.

2 Related Work

This section provides a comprehensive review of automated techniques utilized for the segmentation of Electronic Clinical Narratives (ECNs), comprising both **rule-based** and **machine learning** systems. **Rule-based systems** are grounded on a predetermined set of patterns or rules for discerning section boundaries, either derived by experts or through explicit methods. In contrast, **machine learning** models rely on annotated corpora to train models that can segment and classify sections in new texts. The proposed method for segmenting ECNs, based on a Named Entity Recognition (NER) model, belongs to the **machine learning** classification.

Rule-based methods used for ECN segmentation exhibit varying levels of complexity. Certain studies [12, 13, 14, 15] employed exact matching with tagged headings, while others [16, 16] used regular expressions for section identification in specific document types, such as radiology reports. SecTag [17], a nuanced probabilistic approach employing hierarchical heading terminology and a statistical model for section boundary detection, represents a more advanced methodology that has been adopted in further research [18, 19, 20].

In Electronic Clinical Narratives (ECN), different techniques use **Machine Learning** (ML) to address the identification of sections. For instance, Bramsen et al. [21] identified sections by detecting alterations in temporal focus, while Li et al. [22] approached section mapping as a sequence-labeling problem. Tepper et al. [23] employed an innovative technique that uses BIO tags and category labels to classify text lines in Electronic Clinical Narratives (ECNs). The BIO tag scheme, widely used in Natural Language Processing (NLP), labels tokens as "B" (Beginning), "I" (Inside), or "O" (Outside) based on their relation to a section, offering an effective tool for ECN segmentation. Additional methodologies encompass the use of Support

Vector Machine (SVM) classifiers within the SOAP (Subjective, Objective, Assessment, Plan) framework [24] and Bayesian models for section detection [25]. These diverse ML strategies underscore the versatility and adaptability of machine learning methods for segmenting clinical narratives. Furthermore, the role of annotated datasets turns essential in training models for entity recognition, with semi-automated [13, 14, 26] or active learning techniques [27, 28, 29] being employed to reduce human effort in the annotation process.

The use of NER systems for entity identification in clinical reports has significantly expanded in recent years, primarily due to the advances in deep learning models. The clinical domain has seen the application of numerous strategies, including a two-step method for clinical NER and normalization [30], nested entity extraction from ECNs [31], and the creation of pre-trained language models specifically for NER tasks [32]. These NER models have also been rigorously evaluated under challenging tests to assess their robustness and generalization capabilities [33, 34, 35].

Several corpora have been established for Named Entity Recognition (NER) and normalization tasks within the clinical and biomedical sectors, including but not limited to PharmaCoNER [36], eHealth-KD [37], eHealth CLEF [38], Chilean Waiting List [39], Cantemist [40], and LivingNER [41]. However, ClinAIS uniquely positions itself as the inaugural shared task dedicated explicitly to extracting sections from Electronic Clinical Narratives (ECNs). Previous methods, both **machine learning** and **rule-based**, have limitations, particularly their restriction to identifying entities at the level of words or word groups. To address this, we propose a chunked NER model to identify sections and their corresponding spans within an ECN. However, most of the previous work has been centered on English language texts, thereby creating a gap in resources for languages other than English, such as Spanish. Furthermore, a persistent challenge involves associating each word within an ECN with a corresponding section, given the absence of words in the text without a related section. Its complexity poses a unique obstacle that current methods cannot handle. Our work objective is to reduce these gaps. We present a multidirectional approach. We propose implementing a chunked NER model that identifies entities, sections, and their spans in an ECN, expanding the scope of entity identification beyond individual words or groups of words. Our study specifically concentrates on the task of section classification within Electronic Clinical Narratives (ECNs) in Spanish. Given that most studies within the realm of natural language processing have largely focused on tasks within the English language context, there has been a significant gap in resources and models available for Spanish. Our research, therefore, seeks to address this gap by developing and refining models specifically for section classification in Spanish ECNs. This specialized focus intends to address the unique complexities associated with processing clinical narratives in Spanish. Lastly, we confront the challenge of associating every word within an ECN with a corresponding section, a task that previous methods still need to address adequately. Thus, this work contributes significantly to the field of NER in clinical narratives, pushing the boundaries of current methodologies and addressing previously overlooked challenges.

3   Dataset

The dataset corpus was obtained from the CodiEsp dataset presented in the eHealth CLEF 2020 task [42]. The ClinAIS dataset [10, 11] consists of a collection of 1,038 ECNs annotated with the

beginning and end spans for each of the seven predetermined medical sections:

- **Present Illness:** Outlines the reason for consultation, previous treatments, diagnoses, explorations, and anamnesis.
- **Derived from/to:** Records any patient transfers, including requesting party and reasons.
- **Past Medical History:** Chronicles previous pathologies or notes the absence of such.
- **Family History:** Details family members' pathologies or acknowledges their absence.
- **Exploration:** Covers physical examinations, studies, lab tests, and autopsy findings.
- **Treatment:** Describes treatments or procedures used, including dietary measures.
- **Evolution:** Traces patient's health progression and possible differential diagnoses.

The dataset has three splits: a training set with 781 ECNs (75% of the total), a development set containing 127 ECNs (12.5%), and a test set with 130 ECNs (12.5%). Moreover, the dataset's distribution is stratified by category and annotator, balancing category representation and accounting for different annotator expertise across all subsets.
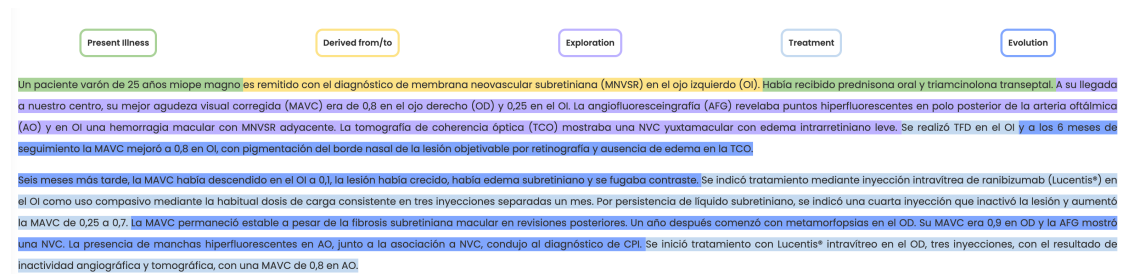

Figure 1: Annotated clinical note example. Source http://ixa2.si.ehu.eus/clinais/taskanddata.

Figure 1 illustrates an example of an annotated Electronic Clinical Narrative (ECN). The image confirms that every word in the ECN belongs to a section. This specific ECN includes five out of the seven possible medical sections. It also shows that sections can appear multiple times and vary in word length.

4   Methodology

This section provides an overview of the proposed model for the ClinAIS shared task on automatic identification of sections in clinical documents.

As shown in Figure 2, we developed a chunked variant of Named Entity Recognition (NER) tailored explicitly for Electronic Clinical Narratives (ECNs). The model can recognize and subsequently categorize distinct sections within an ECN. The categorization process is structured to classify each ECN section into one of the following seven categories: *Present Illness*, *Derived from/to*, *Past Medical History*, *Family History*, *Exploration*, *Treatment*, and *Evolution*.

The model receives as input an Electronic Clinical Narrative (ECN). Then, we employ a text chunking module designed to partition the input data into individual sequences, essentially sectioning the text. This module utilizes a machine learning-based approach, wherein it learns the optimal manner to segregate the data based on training it receives from annotated datasets
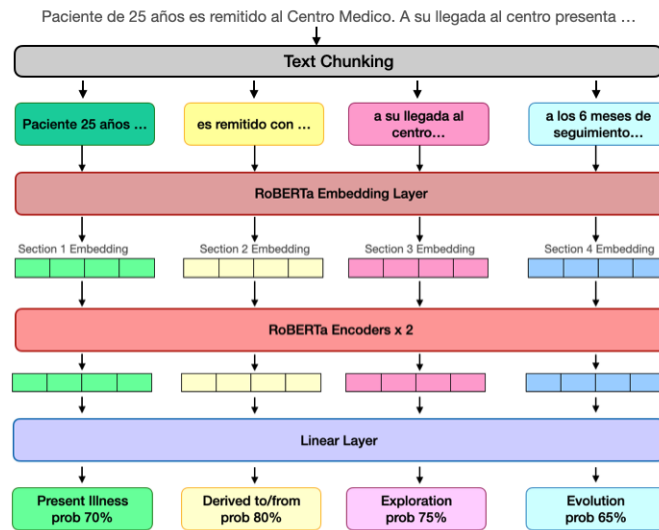
Figure 2: Chunked Named Entity Recognition model architecture.

using BIO tags. It is worth noting that the module facilitates a more granular, context-aware analysis by identifying and separating the text into distinct sections. The efficacy of such a module lies in its ability to adapt and improve its chunking capabilities through iterative learning, thereby becoming more proficient at discerning and distinguishing various sections within a text.

Each section is then converted into an embedding using a RoBERTa model. These embeddings are subsequently processed through a RoBERTa encoder and then passed through an additional RoBERTa encoder for more comprehensive processing. Afterward, the embeddings move through a linear layer. This layer functions to classify each section into the most probable category. It also provides a corresponding probability score, which tells us how confident the model is of each classification. Thus, the final output for each section is a category assignment with a confidence score.

Our methodology employed a RoBERTa model fine-tuned specifically for the Spanish language and Clinical NLP tasks, proposed by Carrino et al. [43]. To adapt the model to our requirements, we set the hidden size to 256, which influences the complexity and capacity of the model. A learning rate of 0.1 was selected to control the model's rate of learning from the data. Additionally, we used a mini-batch size of 4, enabling efficient computation and gradient estimation during the training process. The model was trained over 20 epochs, meaning it went through the entire dataset 20 times to better learn patterns. Importantly, for all layers involved, including both encoders and the embedding layer, we relied on a transformer-based RoBERTa model. This implementation ensured a consistent framework across all processing stages, facilitating a cohesive understanding of the ECN data.

5 Results

The results show the model's performance on the *development set* provided in the ClinIAIS challenge. We evaluated our model using the B2 score [11], a metric designed specifically for the ClinIAIS challenge evaluation. This metric adapts the boundary distance B metric used in text segmentation [44]. The B2 metric incorporates edit distance, including additions, deletions, substitutions, and a transposition operation. This metric is based on borders and boundaries, where a boundary represents the point between two sections. In our evaluation, each token in the note was considered a border. We also calculated Edit Counts, which measure the number of edits required at the boundaries of each ECN to achieve a perfect match with the gold standard and predicted sections. Additionally, we assessed the performance of the proposed model by counting added words, matches, and deletions in each section to check which of them the model performed better.

In the analysis of the results, the B2 score distribution is presented in Figure 3 (a). The majority of the obtained values were above 0.5, indicating a strong tendency towards higher B2 scores. This demonstrates the model's accurately identifying section boundaries within ECNs. Moreover, the B2 score and the distribution obtained in the *development set*, seem to indicate that the model offers the possibility of solving the task slightly differently than the actual annotations. In the face of errors in the annotations, problems arise mainly in some extreme tokens of the annotations due to substitutions, additions, deletions, or transpositions present in the task.
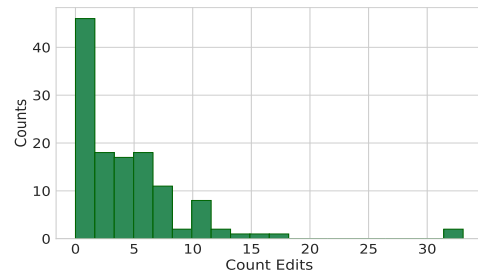
Figure 3 (b) reveals the distribution of the number of edits required for accurately detected sections. Most instances necessitated fewer than five edits per ECN, suggesting a generally robust performance of the model. Specific sections within ECNs, namely *Present Illness*, *Exploration*, and *Treatment*, were identified with exceptional precision by the model, each achieving more than 100 matches, as visualized in Figure 3 (e). Nevertheless, the model faced challenges with specific sections, as illustrated in Figures 3 (c) and (d). The *Exploration* section required the most token additions, amounting to 120, indicating potential under-specification in the model's extraction capabilities. Conversely, the *Present Illness* section necessitated the most token deletions, reaching a total of 40, possibly reflecting over-specification in the model's output. This dissection of the results provides critical insights into the model's strengths and potential areas for improvement. It not only confirms the model's generally sound performance but also points to the necessity for enhanced precision in the detection and extraction of specific sections within ECNs.

6 Limitations

While the model reaches competitive performance in numerous scenarios for section identification in ECNs, it has limitations. A primary limitation comes from the text-chunking module that when not matching the correct section, the error propagates to other sections since, in this task, all parts of the text have an assigned section. This impacts the accuracy of matching all the other sections in the ECN. Another source of error is that since the model performance also depends on the section classification module accuracy, a classification error also negatively affects the general matching accuracy of the model. The effectiveness of the second module is
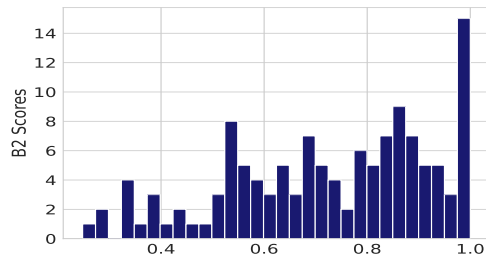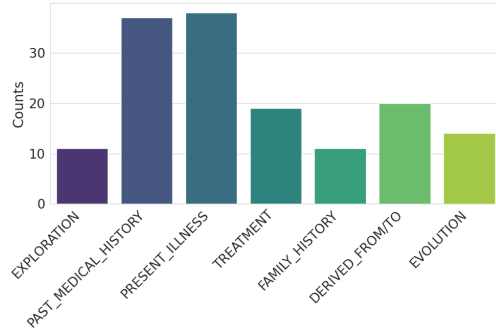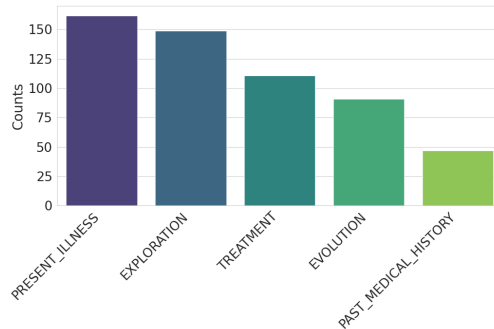
(a) Distribution of B2-score.

(b) Distribution of the number of editions.
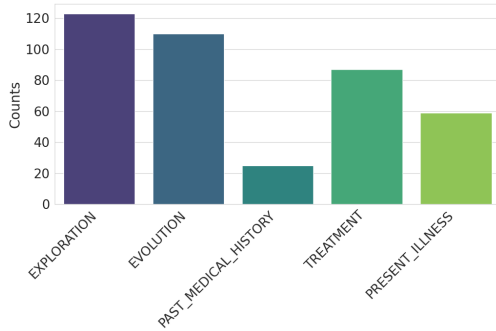
(c) Distribution of the number of added tokens to each section.

(d) Distribution of the number of deleted tokens for each section.

(e) Distribution of the number of matched tokens for each section.

Figure 3: Results on the development set. (a) shows the distribution of B2-score, (b) indicates the distribution of the number of editions, (c) shows the distribution of the number of added tokens to each section, (d) indicates the distribution of the number of deleted tokens for each section, and (e) shows the distribution of the number of matched tokens for each section.

intrinsically linked to the performance of the first module. Addressing these shortcomings could enhance the accuracy and effectiveness of our approach, rendering it a more robust solution for section identification task within ECNs texts.

7 Conclusions and Future Work

In conclusion, our model effectively segments and classifies Spanish Electronic Clinical Narratives (ECNs) sections, exhibiting strong performance in most evaluated sections. However, improvements are needed in token addition and deletion, particularly for the Exploration and Present Illness sections. To address this, we propose two potential avenues for improvement. The first approach involves a two-phase model. In the initial phase, the model would identify section boundaries and accurately segment the text. The second phase would then perform classification within these identified sections, aiming to enhance the precision of predictions. Another potential strategy involves leveraging a model that trains on the current chunked NER output. This model would focus on learning from instances where the predicted spans do not align correctly with the actual sections. By emphasizing the correction of these misaligned predictions, the model could learn to refine its performance and reduce the impact of error propagation. Both of these strategies require further investigation and experimentation to evaluate their effectiveness.

Future research will focus on refining the model to reduce necessary adjustments and improve accuracy across all section types. In this line, we highlight data augmentation by generating paraphrases using generative language models [45]. We can create different variations of the sections for each annotation type and enrich the data set for training in the ClinAIS task. Subsequently, a two-phase model could be evaluated to address the problem of adding and removing tokens in predictions.

References

[1] H. J. Tange, A. Hasman, P. F. de Vries Robbé, H. C. Schouten, Medical narratives in electronic medical records, International journal of medical informatics 46 (1997) 7–29.

[2] A. Pomares-Quimbaya, M. Kreuzthaler, S. Schulz, Current approaches to identify sections within clinical narratives from electronic health records: a systematic review, BMC medical research methodology 19 (2019) 1–20.

[3] E. Iqbal, R. Mallah, R. G. Jackson, M. Ball, Z. M. Ibrahim, M. Broadbent, O. Dzahini, R. Stewart, C. Johnston, R. J. Dobson, Identification of adverse drug events from free text electronic patient records and information in a large mental health case register, PloS one 10 (2015) e0134208.

[4] E. Mahmoudi, N. Kamdar, N. Kim, G. Gonzales, K. Singh, A. K. Waljee, Use of electronic medical records in development and validation of risk prediction models of hospital readmission: systematic review, bmj 369 (2020).

[5] M. M. Paul, C. M. Greene, R. Newton-Dame, L. E. Thorpe, S. E. Perlman, K. H. McVeigh, M. N. Gourevitch, The state of population health surveillance using electronic health records: a narrative review, Population health management 18 (2015) 209–216.

[6] L. Poissant, J. Pereira, R. Tamblyn, Y. Kawasumi, The impact of electronic health records on time efficiency of physicians and nurses: a systematic review, Journal of the American Medical Informatics Association 12 (2005) 505–516.

[7] E. Apostolova, D. S. Channin, D. Demner-Fushman, J. Furst, S. Lytinen, D. Raicu, Automatic segmentation of clinical texts, in: 2009 annual international conference of the IEEE engineering in medicine and biology society, IEEE, 2009, pp. 5905–5908.

[8] J. Demonceau, T. Ruppar, P. Kristanto, D. A. Hughes, E. Fargher, P. Kardas, S. De Geest, F. Dobbels, P. Lewek, J. Urquhart, et al., Identification and assessment of adherence-enhancing interventions in studies assessing medication adherence through electronically compiled drug dosing histories: a systematic literature review and meta-analysis, Drugs 73 (2013) 545–562.

[9] S. M. Jiménez-Zafra, F. Rangel, M. Montes-y Gómez, Overview of IberLEF 2023: Natural Language Processing Challenges for Spanish and other Iberian Languages, in: Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2023), co-located with the 39th Conference of the Spanish Society for Natural Language Processing (SEPLN 2023), CEUR-WS.org, 2023.

[10] I. de la Iglesia, M. Vivó, P. Chocrón, G. de Maeztu, K. Gojenola, A. Atutxa, Overview of ClinAIS at IberLEF 2023: Automatic Identification of Sections in Clinical Documents in Spanish, Procesamiento del Lenguaje Natural 71 (2023).

[11] I. de la Iglesia, M. Vivó, P. Chocrón, G. de Maeztu, K. Gojenola, A. Atutxa, An Open Source Corpus and Automatic Tool for Section Identification in Spanish Health Records, Journal of Biomedical Informatics (2023).

[12] X. Wang, H. Chase, M. Markatou, G. Hripcsak, C. Friedman, Selecting information in electronic health records for knowledge acquisition, Journal of biomedical informatics 43 (2010) 595–601.

[13] T. Edinger, D. Demner-Fushman, A. M. Cohen, S. Bedrick, W. Hersh, Evaluation of clinical text segmentation to facilitate cohort retrieval, in: AMIA Annual Symposium Proceedings, volume 2017, American Medical Informatics Association, 2017, p. 660.

[14] J. Ni, B. Delaney, R. Florian, Fast model adaptation for automated section classification in electronic medical records., MedInfo 216 (2015) 35–9.

[15] R. Waranusast, P. Haddawy, M. Dailey, Segmentation of text and non-text in on-line handwritten patient record based on spatio-temporal analysis, in: Artificial Intelligence in Medicine: 12th Conference on Artificial Intelligence in Medicine, AIME 2009, Verona, Italy, July 18-22, 2009. Proceedings 12, Springer, 2009, pp. 345–354.

[16] R. K. Taira, S. G. Soderland, R. M. Jakobovits, Automatic structuring of radiology free-text reports, Radiographics 21 (2001) 237–245.

[17] J. C. Denny, R. A. Miller, K. B. Johnson, A. Spickard III, Development and evaluation of a clinical note section header terminology, in: AMIA annual symposium proceedings, volume 2008, American Medical Informatics Association, 2008, p. 156.

[18] J. C. Denny, A. Spickard III, P. J. Speltz, R. Porier, D. E. Rosenstiel, J. S. Powers, Using natural language processing to provide personalized learning opportunities from trainee

clinical notes, Journal of biomedical informatics 56 (2015) 292–299.

[19] S. Doan, L. Bastarache, S. Klimkowski, J. C. Denny, H. Xu, Integrating existing natural language processing tools for medication extraction from discharge summaries, Journal of the American Medical Informatics Association 17 (2010) 528–531.

[20] S. Mehrabi, A. Krishnan, A. M. Roch, H. Schmidt, D. Li, J. Kesterson, C. Beesley, P. Dexter, M. Schmidt, M. Palakal, et al., Identification of patients with family history of pancreatic cancer-investigation of an nlp system portability, Studies in health technology and informatics 216 (2015) 604.

[21] P. Bramsen, P. Deshpande, Y. K. Lee, R. Barzilay, Finding temporal order in discharge summaries, in: AMIA annual symposium proceedings, volume 2006, American Medical Informatics Association, 2006, p. 81.

[22] Y. Li, S. Lipsky Gorman, N. Elhadad, Section classification in clinical notes using supervised hidden markov model, in: Proceedings of the 1st ACM international health informatics symposium, 2010, pp. 744–750.

[23] M. Tepper, D. Capurro, F. Xia, L. Vanderwende, M. Yetisgen-Yildiz, Statistical section segmentation in free-text clinical records., in: Lrec, 2012, pp. 2001–2008.

[24] D. Mowery, J. Wiebe, S. Visweswaran, H. Harkema, W. W. Chapman, Building an automated soap classifier for emergency department reports, Journal of biomedical informatics 45 (2012) 71–81.

[25] P. J. Haug, X. Wu, J. P. Ferraro, G. K. Savova, S. M. Huff, C. G. Chute, Developing a section labeler for clinical documents, in: AMIA Annual Symposium Proceedings, volume 2014, American Medical Informatics Association, 2014, p. 636.

[26] A. Carvallo, D. Parra, G. Rada, D. Perez, J. I. Vasquez, C. Vergara, Neural language models for text classification in evidence-based medicine, arXiv preprint arXiv:2012.00584 (2020).

[27] B. Settles, Active learning literature survey (2009).

[28] A. Carvallo, D. Parra, H. Lobel, A. Soto, Automatic document screening of medical literature using word and text embeddings in an active learning setting, Scientometrics 125 (2020) 3047–3084.

[29] A. Carvallo, D. Parra, Comparing word embeddings for document screening based on active learning., in: BIRNDL@ SIGIR, 2019, pp. 100–107.

[30] M. Rojas, J. Barros, M. Araneda, J. Dunstan, Flert-matcher: A two-step approach for clinical named entity recognition and normalization (2022).

[31] P. Báez, F. Bravo-Marquez, J. Dunstan, M. Rojas, F. Villena, Automatic extraction of nested entities in clinical referrals in spanish, ACM Transactions on Computing for Healthcare (HEALTH) 3 (2022) 1–22.

[32] M. Rojas, J. Dunstan, F. Villena, Clinical flair: a pre-trained language model for spanish clinical natural language processing, in: Proceedings of the 4th Clinical Natural Language Processing Workshop, 2022, pp. 87–92.

[33] C. Aspillaga, A. Carvallo, V. Araujo, Stress test evaluation of transformer-based models in natural language understanding tasks, arXiv preprint arXiv:2002.06261 (2020).

[34] V. Araujo, A. Carvallo, C. Aspillaga, D. Parra, On adversarial examples for biomedical nlp tasks, arXiv preprint arXiv:2004.11157 (2020).

[35] V. Araujo, A. Carvallo, C. Aspillaga, C. Thorne, D. Parra, Stress test evaluation of biomedical word embeddings, arXiv preprint arXiv:2107.11652 (2021).

[36] A. Gonzalez-Agirre, M. Marimon, A. Intxaurrondo, O. Rabal, M. Villegas, M. Krallinger, Pharmaconer: Pharmacological substances, compounds and proteins named entity recognition track, in: Proceedings of The 5th Workshop on BioNLP Open Shared Tasks, 2019, pp. 1–10.

[37] L. Monteagudo-Garcıa, A. Marrero-Santos, M. Fernández-Arias, H. Canizares-Dıaz, Uhmmm at ehealth-kd challenge 2021, in: Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2021), 2021.

[38] L. Goeuriot, H. Suominen, L. Kelly, A. Miranda-Escalada, M. Krallinger, Z. Liu, G. Pasi, G. Gonzalez Saez, M. Viviani, C. Xu, Overview of the clef ehealth evaluation lab 2020, in: Experimental IR Meets Multilinguality, Multimodality, and Interaction: 11th International Conference of the CLEF Association, CLEF 2020, Thessaloniki, Greece, September 22–25, 2020, Proceedings 11, Springer, 2020, pp. 255–271.

[39] P. Báez, F. Villena, M. Rojas, M. Durán, J. Dunstan, The chilean waiting list corpus: a new resource for clinical named entity recognition in spanish, in: Proceedings of the 3rd clinical natural language processing workshop, 2020, pp. 291–300.

[40] A. García-Pablos, N. Perez, M. Cuadros, Vicomtech at cantemist 2020, in: Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2020), CEUR Workshop Proceedings, volume 17, 2020, p. 25.

[41] S. Francis, M.-F. Moens, Task-aware contrastive pre-training for spanish named entity recognition in livingner challenge (2022).

[42] A. Miranda-Escalada, A. Gonzalez-Agirre, J. Armengol-Estapé, M. Krallinger, Overview of automatic clinical coding: Annotations, guidelines, and solutions for non-english clinical cases at codiesp track of clef ehealth 2020., CLEF (Working Notes) 2020 (2020).

[43] C. P. Carrino, J. Llop, M. Pàmies, A. Gutiérrez-Fandiño, J. Armengol-Estapé, J. Silveira-Ocampo, A. Valencia, A. Gonzalez-Agirre, M. Villegas, Pretrained biomedical language models for clinical NLP in Spanish, in: Proceedings of the 21st Workshop on Biomedical Language Processing, Association for Computational Linguistics, Dublin, Ireland, 2022, pp. 193–199. URL: https://aclanthology.org/2022.bionlp-1.19. doi:10.18653/v1/2022.bionlp-1.19.

[44] C. Fournier, Evaluating text segmentation using boundary edit distance, in: Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2013, pp. 1702–1712.

[45] Y. Cao, S. Li, Y. Liu, Z. Yan, Y. Dai, P. S. Yu, L. Sun, A comprehensive survey of ai-generated content (aigc): A history of generative ai from gan to chatgpt, arXiv preprint arXiv:2303.04226 (2023).