CBIS'24
XX Congresso Brasileiro de Informática em Saúde
08/10 a 11/10 de 2024 - Belo Horizonte/MG - Brasil

A extração de entidades nomeadas em relatos de casos clínicos

The named entity extraction in clinical case reports

La extracción de entidades nombradas en informes de casos clínicos

Alda Maria Norbiato Torres[1], Raphael Pavani Manhães Bersot[1] , Cristiano da S. Colombo[2]

1 Bacharelando em Sistemas de Informação, Coordenadoria de Sistemas de Informação, Instituto Federal do Espírito Santo, Cachoeiro de Itapemirim (ES), Brasil.
2 Mestre em Cognição e Linguagem, Coordenadoria de Sistemas de Informação, Instituto Federal do Espírito Santo, Cachoeiro de Itapemirim (ES), Brasil.

Autor correspondente: Alda Maria Norbiato Torres
E-mail: aldamntorres@gmail.com
Links: https://github.com/aldatxrres/cbis-ner-spacy

Resumo

É notório que os casos clínicos são utilizados diariamente na rotina de profissionais da saúde, e que podem ser bem aproveitados para originar estudos e criar hipóteses de pesquisas sistematizadas. O presente artigo visa abordar um estudo acerca da extração de informações em relatos de casos clínicos, utilizando a técnica de Reconhecimento de Entidades Nomeadas (REN) para futuro auxílio na investigação de padrões e adversidades em tais relatos. Para o treinamento de uma nova base de conhecimento, foi utilizada a biblioteca spaCy, em Python. Como resultados, foram gerados arquivos HTML com a visualização das entidades reconhecidas e, após os testes, o novo pipeline obteve melhor desempenho ao ser comparado com o modelo pré-treinado nativo do spaCy, atingindo uma acurácia maior que 90% na maior parte dos casos.

Descritores: Reconhecimento de Entidades Nomeadas; Extração de Informações; Relatos de Casos Clínicos

Abstract

It is well-known that clinical cases are used daily in the routine of healthcare professionals and can be effectively utilized to initiate studies and formulate hypotheses for systematic research. This article addresses a study on information extraction from clinical case reports, employing the Named Entity Recognition (NER) technique to aid in investigating patterns and adversities in such reports. The spaCy library in Python was employed to train a new knowledge base. As a result, HTML files were generated for the visualization of recognized entities, and after testing, the new pipeline showed superior performance compared to the native pre-trained spaCy model, achieving an accuracy greater than 90% in most cases.

Keywords: Named Entity Recognition; Information extraction; Clinical Case Reports

Resumen

Es bien sabido que los casos clínicos se utilizan a diario en la rutina de los profesionales de la salud y pueden ser aprovechados eficazmente para iniciar estudios y formular hipótesis para investigaciones sistemáticas. Este artículo tiene como objetivo abordar un estudio sobre la extracción de información de informes de casos clínicos, empleando la técnica de Reconocimiento de Entidades Nombradas (REN) para ayudar en la investigación de patrones y adversidades en dichos informes. Para el entrenamiento de una nueva base de conocimientos, se utilizó la biblioteca spaCy en Python. Como resultados, se generaron archivos HTML para la visualización de entidades reconocidas y, después de las pruebas, el nuevo pipeline mostró un rendimiento superior en comparación con el modelo pre-entrenado nativo de spaCy, alcanzando una precisión mayor del 90% en la mayoría de los casos.

Descriptores: Reconocimiento de Entidades Nombradas; Extracción de Información; Informes de casos clínicos

Introdução

Os relatos de casos são uma forma de compartilhar informações cruciais para a prática médica. Além disso, eles desempenham um papel significativo como recursos educacionais valiosos tanto para os autores quanto para os leitores [1]. Ao analisar esses textos médicos, observa-se que eles constituem valiosas fontes de informações, fornecendo *insights* essenciais sobre o estado de saúde dos pacientes.

No entanto, pode ser um desafio extrair informações de um grande volume de relatos clínicos realizando a leitura manual de cada documento, visto que grande parte dos registros médicos são feitos por meio de linguagem natural não estruturada[2]. Considerando a origem heterogênea, semiestruturada ou não estruturada, a interpretação e correlação de relatos de casos clínicos também se torna um desafio notável[3].

Nesse contexto, a abordagem de processamento computacional, com ênfase no Processamento de Linguagem Natural (PLN), surge como uma solução promissora para realizar a extração de informações dos casos clínicos. Para a construção do experimento deste artigo, foi utilizada uma biblioteca gratuita e de código aberto em Python, denominada spaCy. Esta biblioteca foi desenvolvida com o propósito de facilitar a construção de aplicações voltadas para o processamento e compreensão de grandes volumes de texto.

Neste trabalho é apresentado um modelo treinado em Língua Portuguesa para a extração das Entidades Nomeadas (ENs). A categorização e a extração de informações podem auxiliar os profissionais da saúde na tomada de decisão, melhorando a qualidade do trabalho e favorecendo o atendimento dos pacientes.

Este artigo, portanto, investiga e utiliza técnicas de extração de informações dos casos clínicos, em particular, a técnica de Reconhecimento de Entidades Nomeadas (REN).

Processamento de Linguagem Natural

O Processamento de Linguagem Natural (PLN) é um campo de pesquisa que tem como objetivo investigar e propor métodos e sistemas de processamento computacional da linguagem natural[4]. Refere-se por linguagem natural uma linguagem que é usada para comunicações do dia-a-dia feitas por humanos; línguas como Português, Inglês ou Mandarim.[5]. As tecnologias de PLN capacitam os sistemas computacionais a realizar de

forma automatizada a extração de informações e conhecimentos a partir de dados não estruturados da linguagem humana[6].

No contexto deste trabalho, o PLN possibilita a extração de informações valiosas a partir de relatos de casos clínicos redigidos em linguagem natural não estruturada, envolvendo a identificação de padrões, sintomas, doenças e outros indicadores relevantes para possíveis diagnósticos médicos.

Reconhecimento de Entidades Nomeadas

As Entidades Nomeadas (EN) são como elementos textuais encontrados em documentos escritos em linguagem natural[7].

> Essas entidades podem ser categorizadas em tipos pré-definidos, como Pessoa, Organização, Local, Data, Moeda, entre outros, sendo denominadas como entidades pré-construídas. Além disso, as entidades podem receber classificações específicas de acordo com a aplicação em questão, exigindo a aplicação de algoritmos de aprendizado de máquina para o treinamento prévio dessas entidades [7].

Uma técnica para realizar esta tarefa é o Reconhecimento de Entidades Nomeadas (REN), que "define-se como uma tarefa cujo objetivo é identificar as entidades nomeadas bem como sua posterior classificação, atribuindo uma categoria semântica para essas entidades[8]". Essa tarefa envolve identificar nomes próprios em um texto específico e categorizá-los em diferentes categorias de interesse ou em uma categoria padrão denominada "Outros"[9].

Um exemplo de aplicação desta técnica em textos biomédicos, é a extração de informações de bulas de medicamentos[10]. Neste trabalho, os autores extraíram informações de bulas de medicamentos relacionadas a nomes de medicamentos, princípios ativos, doenças, sintomas de doenças e pessoas.

Relatos de Casos clínicos

Um relato de caso é uma narrativa abrangente que fornece uma descrição clara e detalhada de experiências médicas únicas com pacientes, as quais podem ter impacto tanto nas práticas clínicas quanto nas de pesquisa. Este tipo de texto contribui para o

aumento do conhecimento existente sobre importantes tópicos clínicos e para fornecer percepções sobre doenças novas ou raras e tratamentos não convencionais, podendo posteriormente ser avaliadas de modo mais formal utilizando desenhos de estudo mais sofisticados, como ensaios controlados randomizados[13].

Além disso, estes textos apresentam casos com descrições originais de observações clínicas. Os relatos devem representar originalidade quanto à abordagem do diagnóstico, tratamento ou resultados. Os autores devem apresentar, assim, situações raras ou pouco frequentes no ambiente clínico, justificando atenção e uma análise detalhada sobre o tema[12].

Um exemplo de trecho de caso clínico utilizado, é apresentado a seguir:

"Sexo feminino, 73 anos, leucodérmica, com antecedentes pessoais conhecidos de hipertensão arterial, diabetes mellitus tipo 2 insulinotratada, dislipidemia e doença cerebrovascular. Doente negou alergias medicamentosas conhecidas."[18]

A Biblioteca spaCy

De acordo com a documentação oficial, spaCy é uma biblioteca de processamento de linguagem natural avançada e de código aberto desenvolvida em Python[11]. Seu principal propósito é fornecer uma ferramenta robusta para o processamento eficiente de grandes volumes de texto, podendo ser utilizado para construir sistemas de extração de informações ou de compreensão de linguagem natural. spaCy oferece modelos pré-treinados para diferentes idiomas, permitindo aos usuários realizar tarefas de PLN com facilidade em várias línguas[11].

Trabalhos Correlatos

Nesta seção são apresentados trabalhos correlatos à proposta desta pesquisa.

Um trabalho apresentou etiquetadores desenvolvidos para textos em português, refinados a partir dos modelos BioBERTpt (clínico/biomédico) e BERTimbau (genérico) em um corpus com anotações morfossintáticas[17]. Segundo os autores, foi obtido 0.9826 de acurácia, estado-da-arte para o corpus utilizado. Também foi realizada uma avaliação por humanos dos modelos treinados e outros da literatura, utilizando narrativas clínicas autênticas. O modelo clínico atingiu 0.8145 de acurácia comparado com 0.7656 do modelo genérico.

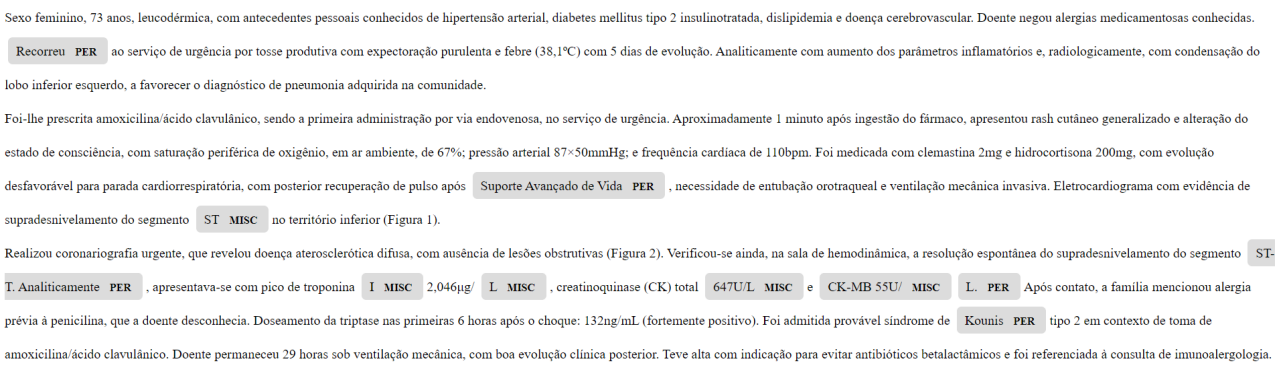
Um estudo para identificar automaticamente as entidades relacionadas às reações adversas a medicamentos a partir das descrições narrativas dos Relatórios Chineses de Eventos Adversos a Medicamentos, propõe que tais informações sejam usadas de forma complementar na avaliação da seção estruturada de casos, podendo auxiliar ainda mais na avaliação de Reações Adversas a Medicamentos[16]. Neste estudo foram utilizados dois modelos REN de CRF e BiLSTM-CRF, bem como um modelo BiLSTM-CRF baseado em recursos lexicais (LF-BiLSTM-CRF) gerado para realizar tarefas REN. Entre os três modelos, o LF-BiLSTM-CRF obteve a maior pontuação média na F1 de 94,35%.

Uma rede neural composta por BiLSTM com uma camada sequencial CRF onde diferentes *word embeddings* são combinados como entrada para a arquitetura, de modo que um método híbrido combinando modelos supervisionados e não supervisionados é usado para a tarefa de indexação de conceitos[15]. Esta rede neural foi utilizada no reconhecimento de substâncias químicas e medicamentos, além de indexar as entidades usando a terminologia SNOMED-CT (terminologia de referência no domínio biomédico que permite atribuir um código identificador único a cada entidade reconhecida). Este trabalho apresentou o resultado na medida F1 foi de 90,77% para o reconhecimento de medicamentos e substâncias químicas.

Metodologia de Trabalho

Ao realizar os testes com nossa base de relatos de casos com a biblioteca spaCy, utilizando o *pipeline* (neste caso, é basicamente um pacote de dados treinados) em Português disponibilizado nativamente, os resultados obtidos foram mínimos e insuficientes para o caso específico da área da saúde. Com isso, não reconhecendo medicamentos, sintomas ou qualquer outra entidade que fora julgada importante para esse contexto. Isto se justifica pelo fato dos modelos pré-treinados não serem contemplados com textos deste domínio específico, ou seja, da área de saúde. A Figura 1 apresenta os resultados obtidos nesta etapa do trabalho.

Figura 1 – Visualização de um caso clínico com as entidades reconhecidas com spaCy



Com isso, observou-se a possibilidade de uso de outra biblioteca, baseada no spaCy, porém voltada para textos médicos, a MedspaCy. No que tange a biblioteca MedspaCy, não obtivemos nenhum resultado ao processar os mesmos relatos de caso usados durante os testes com spaCy. Isso aconteceu porque essa biblioteca não possui nenhuma base de dados disponível em Língua Portuguesa e, levando em consideração o cenário deste trabalho, fez-se necessário o desenvolvimento e treinamento de um modelo próprio, utilizando como base o spaCy, que contemplasse os devidos termos da área da saúde para o reconhecimento das entidades desejadas em Língua Portuguesa.

Durante o treinamento, foram definidas 7 categorias para a classificação das entidades, sendo elas: procedimento, doença, medicamento, sintoma, especialidade, reação e diagnóstico, escritas em maiúsculo e sem nenhum acento ou caractere especial. Tais etiquetas foram pensadas levando em consideração o conteúdo dos casos clínicos utilizados nos experimentos, sendo suficientes para catalogar as informações que eram coerentes para o cenário.

A anotação dos relatos de casos foi realizada por uma equipe de 3 pessoas, sendo uma delas com conhecimentos na área da saúde. As anotações utilizaram os conceitos descritos no Código Internacional de Doenças (CID-10) e também no Dicionário de Especialidades Farmacêuticas (DEF).

Foram utilizados 10 arquivos no formato .txt, onde 6 deles eram devidamente anotados na formatação suportada pelo spaCy, para a realização do treinamento, e 4 arquivos não anotados para os testes posteriores. O padrão de anotação suportado pela biblioteca seguiu uma estrutura semelhante ao de um dicionário em Python, onde eram descritas as etiquetas dentro de um objeto *"entities"* e, para cada etiqueta identificada no texto, foi atribuída o número do caractere inicial onde ela se encontrava e o número do caractere final. Este trabalho de anotação foi todo feito manualmente pelos autores, arquivo por arquivo. A Figura 2 mostra o exemplo de um trecho anotado.

Figura 2 – Exemplo de anotação de um relato de caso

```
("Sexo feminino, 73 anos, leucodérmica, com antecedentes
pessoais conhecidos de hipertensão arterial, diabetes mellitus
tipo 2 insulinotratada, dislipidemia e doença cerebrovascular.
Doente negou alergias medicamentosas conhecidas.", {"entities":
[(15, 22, "IDADE"), (24, 36, "DOENCA"), (78, 98, "DOENCA"),
(100, 124, "DOENCA"), (125, 140, "MEDICAMENTO"), (142, 154,
"DOENCA"), (157, 179, "DOENCA")]})
```

Para uma melhor visualização dos resultados e da classificação de cada etiqueta, foi utilizada outra biblioteca, o displaCy, que permitiu diferenciar as etiquetas com um padrão de cores, que pode ser observado na Figura 3. Ao executar o código, o displaCy é responsável por criar um arquivo HTML com as entidades reconhecidas.

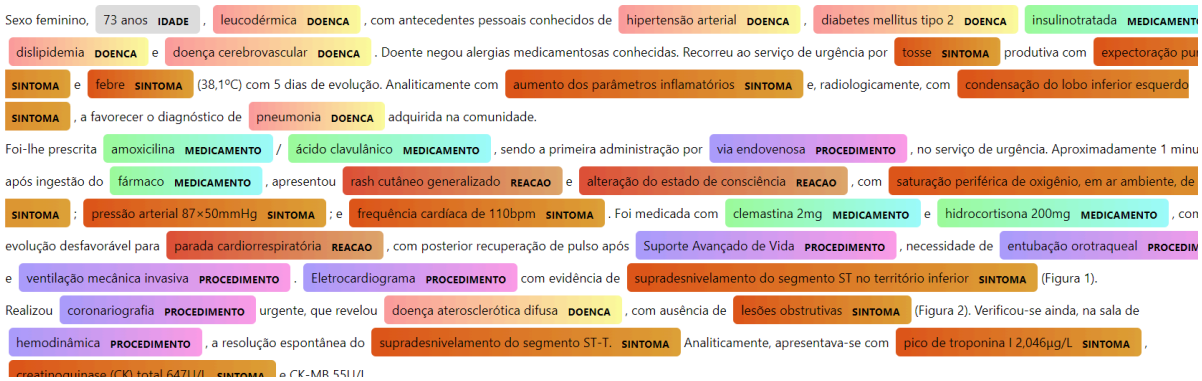
As entidades que aparecem na cor cinza nos resultados foram reconhecidas de forma nativa pelo modelo base do spaCy. O *script* em Python usado para criar e treinar um novo *pipeline* foi desenvolvido com base na documentação original disponibilizada no *website* da biblioteca e no conhecimento prévio dos autores diante da linguagem de programação em questão.

Os testes foram executados utilizando os 10 relatos de casos clínicos, incluindo os que foram anotados anteriormente e usados para o treinamento do *pipeline*, o que contribuiu para a realização da comparação de desempenho entre os dados anotados e não anotados. Porém, no treinamento, os arquivos anotados previamente foram inseridos como textos puros, sem anotação.

Experimentos e Resultados

Foram realizados dois experimentos: o primeiro testou o modelo criado no conjunto de documentos utilizado em seu treinamento; e o segundo experimento testou o modelo num conjunto de relatos de casos não utilizado em seu treinamento.

Figura 3 – Visualização de parte de um caso clínico pós treinamento



Os resultados do primeiro experimento demonstraram que o modelo foi capaz de reconhecer com exatidão todas as entidades presentes nos textos do treino, conforme mostra a Tabela 1. Isto demonstrou que o processo de anotação dos documentos para criar uma base de treino e o treinamento do modelo a partir desta base gerou um modelo capaz de reconhecer entidades da área de saúde nos relatos de casos clínicos. Caso o modelo não reconhecesse as entidades presentes nestes documentos, o processo seria revisado e refeito.

Tabela 1 – Comparativo de resultados casos clínicos anotados e usados em treinamento

Caso clínico	Qtd. ent. reconhecidas	% ent. reconhecidas	Qtd. ent. corretas	% ent. corretas
cc_010[18]	45	100	45	100
cc_021[19]	30	100	29	96,66
cc_032[20]	26	100	26	100
cc_034[21]	52	100	52	100
cc_035[21]	26	100	22	84,61

Os resultados do segundo experimento apontam que o modelo foi capaz de reconhecer um padrão entre algumas entidades. Ele identificou mais de 60% das entidades de maneira correta em todos os documentos de teste, conforme apresentado na Tabela 2. Para o preenchimento de dados das tabelas, foram considerados o número de entidades reconhecidas pelo modelo em cada arquivo HTML processado e, dentre

essas entidades reconhecidas, quais delas estavam realmente corretas, de acordo com os critérios de anotação.

Tabela 2 – Comparativo de resultados casos clínicos não anotados e usados em treinamento

Caso clínico	Qtd. ent. reconhecidas	% ent. reconhecidas	Qtd. ent. corretas	% ent. corretas
cc_036[21]	5	100	3	60
cc_037[21]	20	100	16	80
cc_038[21]	19	100	14	73,68
cc_039[21]	14	100	9	64,28

Para se verificar de forma mais efetiva a qualidade do modelo criado, os resultados apresentados na Tabela 3 são expressos através de *Precision* (P), *Recall* (R) e *F-measure* (F1). É possível verificar o quanto o algoritmo foi capaz de alcançar um equilíbrio entre cometer menos erros enquanto classifica corretamente cada entidade reconhecida[14].

Tabela 3 – Resultados do experimento de teste em medida F1.

Categorias	P	R	F1
Doenca	0,0	0,0	0,0
Diagnostico	0,0	0,0	0,0
Especialidade	0,0	0,0	0,0
Medicamento	80,0	100,0	89,0
Procedimento	0,0	0,0	0,0
Reacao	43,0	100,0	60,0
Sintoma	40,0	100,0	57,0

É possível perceber que o modelo não reconheceu nenhuma entidade das categorias Doenca, Diagnostico, Especialidade e Procedimento. Isto demonstra que há necessidade de melhorar a qualidade do treino para que o modelo seja capaz de reconhecer entidades destas categorias.

Por outro lado, percebe-se que o modelo reconheceu entidades das categorias

Medicamento, Reacao e Sintoma. Ou seja, mesmo com uma quantidade pequena de documentos no treino o modelo obteve taxa de assertividade razoável para entidades das categorias Reacao (60,0) e Sintoma (57,0) em medida F1. Isso pode ser justificado pela característica dos textos utilizados no treinamento. A taxa de acerto para as entidades da categoria Medicamento (89,0) em medida F1 foi superior, indicando o potencial do modelo.

Conclusões

Este trabalho apresentou um método para treinamento e teste de um modelo capaz de extrair informações de relatos de casos clínicos, através do uso de técnicas de PLN e da biblioteca spaCy em Língua Portuguesa. Ao combinar uma metodologia de anotação cuidadosa com o treinamento de um modelo específico para a área da saúde, obtivemos resultados satisfatórios.

Mesmo com um conjunto de treinamento reduzido e um ambiente criado para fins de pesquisa, os resultados obtidos mostraram que o modelo treinado superou significativamente o desempenho do modelo base do spaCy, reconhecendo entidades que não eram identificadas nativamente. Desta forma, propõe-se que a assertividade do modelo seja aprimorada se a equipe de anotação fosse composta por mais profissionais da saúde e especialistas, atuando na revisão das anotações e iterando a base a cada melhoria. Além disso, a utilização de um método semi-automático na etapa de anotação, traria mais agilidade ao processo, deixando-o mais rápido e menos suscetível a erros.

É importante destacar o esforço deste trabalho para demonstrar que a adaptação de ferramentas de PLN para domínios específicos é possível e viável conforme demonstraram os resultados. Nesta perspectiva, como trabalho futuro pretende-se aprimorar o modelo criado com o uso de um número maior de documentos na fase de treinamento, bem como ampliar o escopo de testes deste modelo.

A extração de informações de relatos de casos clínicos demonstrou que a iniciativa deste trabalho pode ser direcionada futuramente para prover *insights* que sejam úteis como suporte aos profissionais da área da saúde em tomadas de decisões ou para melhorar a análise de dados clínicos. Tal afirmativa se baseia nos resultados obtidos pelo

modelo, os quais indicam que a extração de entidades relacionadas às doenças e aos medicamentos pode auxiliar profissionais de saúde e gestores na aquisição de conhecimentos para tomadas de decisões, bem como na investigação sobre interações medicamentosas e possíveis reações adversas ou na realização de pesquisas sobre novos tratamentos terapêuticos[22]. Nesta mesma linha de aplicação, é possível apontar que a extração das entidades nomeadas de textos da área de saúde, pode ser usada por um Sistema de Informação Inteligente para sanar dúvidas de pacientes e auxiliar profissionais de saúde na tomada de decisão sobre terapias e prescrição de medicamentos[10]. Além disso, a extração destas entidades pode se tornar uma rica ferramenta de apoio na produção de material didático e pedagógico para educadores que atuam em cursos da área de saúde, visto que é possível a geração automática de perguntas e respostas a partir de um modelo REN eficiente[23]. Os exemplos descritos e citados demonstram como o modelo criado e seus resultados podem ser utilizados para auxiliar de forma positiva o dia a dia dos profissionais de saúde.

Futuramente, pretende-se inserir outros tipos de textos da área de saúde, tais como bulas de medicamentos, prontuários médicos, etc…, nas fases de treino e teste do modelo. Além disso, pretende-se ampliar o escopo de nosso trabalho com o intuito de extrair as relações entre as entidades reconhecidas nestes tipos de documentos.

Prover uma colaboração entre profissionais da saúde e especialistas em PLN é crucial para garantir a precisão e relevância dos resultados. Este trabalho é um passo na direção de sistemas REN robustos e precisos para a área da saúde em Português.

Referências

1. CARLETON HA, Webb ML. The case report in context. Yale J Biol Med. 2012;85(1):93-96.

2. SANTOS, DB. Visualização de dados estruturados e não estruturados da área da saúde. Universidade Estadual Paulista (Unesp), 2022.

3. RIEDO, SAC. Vitor dos S. Um modelo para extração, estruturação, indexação e recuperação de casos clínicos publicados na web. ISSN 2178-8332.

4. CASELI, HM; NUNES, MGV. Processamento de linguagem natural: conceitos, técnicas e aplicações em português. 2023.

5. BARBOSA, J. et al. Introdução ao processamento de linguagem natural usando python. III Escola Regional de Informática do Piauí, v. 1, p. 336–360, 2017.

6. DE JESUS FALCÃO, LC; LOPES, B; SOUZA, RR. Absorção das tarefas de processamento de Linguagem Natural (NLP) pela Ciência da Informação (CI): uma revisão da literatura para tangibilização do uso de NLP pela CI. Em Questão, p. 13-34, 2022.

7. KIRSCH, BG; DORNELES, ÁP. Desenvolvimento de uma ferramenta para reconhecimento de entidades nomeadas em certificados de atividades complementares de curso utilizando spacy. Anais do Encontro Anual de Tecnologia da Informação, v. 12, n. 1, p. 44–44, 2023.

8. AMARAL, DOF. O reconhecimento de entidades nomeadas por meio de conditional random fields para a língua portuguesa. 2013. Dissertação de Mestrado. Pontifícia Universidade Católica do Rio Grande do Sul.

9. MILIDIú RUY LUIZ DUARTE, JCCR. Machine learning algorithms for Portuguese named entity recognition. Inteligência Artificial. Revista Iberoamericana de Inteligencia Artificial, 2007. ISSN 1137-3601. Disponível em: https://www.redalyc.org/articulo.oa?id=92503610

10. COLOMBO, CS; OLIVEIRA, ES. Intelligent information system for extracting knowledge from pharmaceutical package inserts. In: Proceedings of the XVIII Brazilian Symposium on Information Systems. [S.l.: s.n.], 2022. p. 1–9

11. SPACY. spaCy 101: Everything you need to know · spaCy Usage Documentation. Disponível em: <https://spacy.io/usage/spacy-101>.

12. Revista PubSaúde. Relato de Caso Clínico. Disponível em: https://pubsaude.com.br/artigo-original/relato-de-caso-clinico/. Acesso em: 13 mar. 2024. [2024?].

13. PATINO, CM, FERREIRA, JC: Relatos de caso: narrativas destacando experiências clínicas que contribuem para a prática e para futuros estudos. Jornal Brasileiro de Pneumologia. 45, (2019). https://doi.org/10.1590/1806-3713/e20190251.

14. SANTOS, HDP, ULBRICH, AHDPS, VIEIRA, R. Evaluation of a Prescription Outlier Detection System in Hospital's Pharmacy Services, Anais do IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2862–2868, 7, (2021).

15. LÓPEZ-ÚBEDA, P, DÍAZ-GALIANO, MC, UREÑA-LÓPEZ, A, MARTIN-VALDIVIA, MT: Combining word embeddings to extract chemical and drug entities in biomedical literature. BMC Bioinformatics 22(1), 1–17 (2021).

16. YAO CHEN, CZ, Tianxin Li, HW, Kai Ye, XZ, Jun, L. 2019. Named entity recognition from Chinese adverse drug event reports with lexical feature based BiLSTM-CRF and tri-training. Journal of Biomedical Informatics 96, 1 (jul 2019), 1–9. https://doi.org/10.1016/j.jbi.2019.103252.

17. SCHNEIDER, ETR, GUMIEL, YB, OLIVEIRA, LFA de, MONTENEGRO, CO, BARZOTTO, LR, MORO, C, Paraiso, EC. (2023). Developing a Transformer-based Clinical Part-of-Speech Tagger for Brazilian Portuguese. Journal of Health Informatics, 15(Especial). https://doi.org/10.59681/2175-4411.v15.iEspecial.2023.1086.

18. DUARTE, P, et al. Síndrome de Kounis: A propósito de um clínico. Revista Brasileira de Terapia Intensiva. 2020;32(1):149-152. https://doi.org/10.5935/0103-507X.20200021.

19. FILHO, ESM, et al. Púrpura trombocitopênica trombótica associada à gravidez: Relato de caso. Rev Bras Ter Intensiva. 2009; 21(3):336-339. https://doi.org/10.1590/S0103-507X2009000300016.

20. VIEIRA, DF, et al. Interação dos medicamentos glibenclamida e furosemida em paciente com hipertensão e diabetes: Relato e estudo de caso clínico. Enciclopédia Biosfera, 8(14). https://conhecer.org.br/ojs/index.php/biosfera/article/view/3978.

21. PSIQUIATRIA GERAL. Casos Farmacológicos. Disponível em: <https://www.psiquiatriageral.com.br/tratamento/interacoes14.htm>. Acesso em 17/02/2022.

22. COLOMBO, CS; OLIVEIRA, ES. A Extração de Entidades Nomeadas em Bulas de Medicamentos e em Relatos de Casos Clínicos. In: Anais do XXIV Simpósio Brasileiro de Computação Aplicada à Saúde. [S.l.: s.n.], 2024. p. 627–638. https://doi.org/10.5753/sbcas.2024.

23. PIROVANI, JPC; SPALENZA, MA; OLIVEIRA, E. Geração Automática de Questões a Partir do Reconhecimento de Entidades Nomeadas em Textos Didáticos. In: Anais do XXVIII Simpósio Brasileiro de Informática na Educação, 2017. p. 1147–1156.