



MODALIDADE: Comunicação oral

EIXO TEMÁTICO: Tecnologias disruptivas e a Arquivologia

**Extração de Metadados para um Grande Arquivo de
Decisões Judiciais: Uma Abordagem com Inteligência
Artificial**

*Metadata Extraction for a Large Archive of Judicial
Decisions: An Artificial Intelligence Approach*

**João Lima², Aerty Santos⁴, Eduardo Almeida⁴, Juliana
Pirovani^{3,5}, Elias de Oliveira^{1,3,4}**

¹Departamento de Arquivologia, ²Departamento de Informática,

³Universidade Federal do Espírito Santo, ⁴Programa de Pós-

Graduação em Informática, ⁵Departamento de

Computação/Alegre

Resumo

A descrição arquivística costuma ser uma tarefa manual e demorada. Dessa forma, tende a ser mais estática em seu ciclo de vida da informação. Por outro lado, os usuários mudam com maior frequência a forma com que utilizam os documentos e exigem novos recursos não previstos na descrição inicial dos documentos no arquivo. Esse artigo apresenta uma estratégia para encontrar entidades nomeadas automaticamente, de forma a produzir elementos ricos de metadescrição para um arquivo a ser gerido no AtoM. Para atingir esse objetivo, aplicamos algumas estratégias de inteligência artificial. Em nossos experimentos, usamos o conjunto de dados de decisões do Tribunal de Justiça de São Paulo em um total de 9.178 e realizamos a extração de alguns novos elementos descritivos. A partir deste corpus, separamos um conjunto menor de documentos relacionados as decisões judiciais dos crimes de Roubo e Furto. Os experimentos demonstraram que a abordagem proposta reduziu significativamente o esforço humano necessário para a realização da tarefa, utilizando a metodologia baseada em inteligência artificial.

Palavras-chave: AtoM; decisões judiciais; inteligência artificial.

Abstract

Archival description is usually a manual and time-consuming task. As a result, it tends to be more static



in its information lifecycle. On the other hand, users more frequently change the way they use documents and demand new features not foreseen in the initial description of the documents in the archive. This article presents a strategy for automatically finding named entities, in order to produce rich meta-description elements for an archive to be managed in AtoM. To achieve this objective, we applied some artificial intelligence strategies. In our experiments, we used the dataset of decisions from the São Paulo Court of Justice, totaling 9,178 documents, and extracted some new descriptive elements. From this corpus, we separated a smaller set of documents related to judicial decisions on Robbery and Theft crimes. The experiments demonstrated that the proposed approach significantly reduced the human effort required to complete the task, utilizing the methodology based on artificial intelligence.

Keywords: AtoM; judicial decisions; artificial intelligence.

1 INTRODUÇÃO

As ferramentas de recuperação de informações melhoraram na última década. Google, Yahoo, Microsoft e outras empresas oferecem mecanismos de busca poderosos que permitem pesquisar mundialmente por grandes quantidades de documentos quase instantaneamente. Infelizmente, ferramentas semelhantes ainda não estão disponíveis em muitos sistemas de informação usados pelos departamentos de justiça no Brasil. Quando integradas a um sistema de informação institucional local, ferramentas como essas podem aumentar a eficiência e precisão na busca por informações em grandes arquivos de documentos (Izo; Oliveira; Badue, 2021).

A decisão do juiz é o ato administrativo. Um marco emitido pelo sistema judicial referente a um fato disputado entre partes. Tendo isso em mente, esses documentos são valiosos instrumentos para avaliar o desempenho do sistema judicial. No entanto, para atingir esse objetivo, precisamos construir um melhor sistema de arquivamento para aumentar a acessibilidade desses documentos, vinculando-os a novos pontos de acesso. Os acessos atuais são baseados principalmente no número do processo, no nome das partes e no tipo de assunto em disputa, por exemplo.

A extração de alguns desses metadados pode ser feita diretamente por meio de estratégias baseadas no reconhecimento de padrões. Neste trabalho, vamos além desse tipo de procedimento, extraindo outras informações presentes nesses documentos: por exemplo, o nome das pessoas mencionadas, datas dos fatos mencionados e o tempo aplicado à sentença, apenas para citar algumas dessas informações contidas no texto da decisão. Para peças de informação mais complexas, são necessárias técnicas de Processamento de Linguagem Natural (Manning; Schuetze, 1999). Nossa abordagem segue aquela apresentada por (Pirovani; Oliveira, 2021). A abordagem daqueles autores usa uma Gramática Local (LG)



para cada entidade de interesse de anotação, auxiliando o Conditional Random Field (CRF) (Sutton; McCallum, 2012), outro algoritmo inteligente. O papel de uma LG é inserir o conhecimento do especialista do domínio no processo, já que uma LG é construída de forma eminentemente manual. Algumas abordagens alternativas são vorazes em dados para treinar seus algoritmos (Santos; Guimaraes, 2015), o que pode ser proibitivo em determinado contexto, ou devido ao limitado orçamento da equipe responsável.

Este trabalho oferece duas contribuições principais. A primeira é a disponibilização pública de um extenso conjunto de decisões judiciais relacionadas aos crimes abordados. A segunda contribuição reside na organização dessas decisões de maneira compatível com o AtoM¹.

O AtoM é uma das duas plataformas recomendadas pelo RDC-Arq (Repositórios Arquivísticos Digitais Confiáveis) para a gestão e descrição de documentos arquivísticos, sendo reconhecido por suas robustas funcionalidades em organização, difusão, transparência ativa e compartilhamento de informações arquivísticas. A plataforma opera de forma eficiente e em conformidade com os padrões internacionais. Enquanto isso, a plataforma Archivematica, também recomendada pelo RDC-Arq, foca na preservação digital, assegurando a integridade e autenticidade dos documentos ao longo do tempo. (CONARQ, 2023).

Para apresentar nossa proposta, consideramos dois tipos de crimes com decisões proferidas em 2022: furto e roubo. Utilizamos 5.025 documentos em nossos experimentos. Este artigo está estruturado da seguinte forma: na Seção 2, discutimos trabalhos recentes que servirão de base para comparar nossa proposta. Na Seção 3, abordamos a busca por estruturas textuais mais complexas para a criação de metadados inteligentes, como apontado por Izo, Oliveira e Badue (2021), visando facilitar o acesso dos usuários. A Seção 4 detalha os resultados dos experimentos, que validam nossa hipótese. Nossa pesquisa expandiu a gama de metadados extraídos, resultando em uma redução significativa na carga de trabalho dos especialistas humanos em comparação aos processos manuais. Por fim, na Seção 5, apresentamos nossas conclusões e propostas para pesquisas futuras.

2 REVISÃO DE LITERATURA

Um estudo realizado por Moraes, Zafalon e Barroso (2019) examinou a relação entre os

¹<https://www.accesstomemory.org/pt-br/>



avanços tecnológicos e a reconfiguração dos sistemas de informação na descrição arquivística. A pesquisa teve como objetivo investigar a produção científica relacionada ao modelo conceitual internacional de descrição arquivística, *Records in Contexts* (RiC), e ao software AtoM, ambos desenvolvidos pelo *International Council on Archives* (ICA). Utilizando uma abordagem que combina métodos qualitativos e quantitativos, incluindo revisão bibliográfica e análise documental, o estudo identificou tendências e analisou como a literatura discute a aplicação dessas ferramentas. A pesquisa revelou que o modelo RiC e o software AtoM são considerados fundamentais para a organização e recuperação de documentos arquivísticos, promovendo a interoperabilidade entre sistemas e instituições.

Outro estudo, conduzido por Campos e Oliveira (2015), propôs uma metodologia baseada em gramática local para a identificação automática de nomes de pessoas em textos em português, usando como exemplo o romance *Senhora*, de José de Alencar. A estratégia adotada, que integra informações contextuais e dicionários, resultou em indicadores de desempenho notáveis, alcançando 99% de recall e 100% de precisão na identificação de 1699 nomes quando aplicada ao livro *Senhora*. No entanto, ao estender essa abordagem para a classe Tudo a ver (TAV) do conjunto de dados ATribuna, apenas 63 dos 116 nomes existentes foram corretamente identificados, com 9 falsos positivos, resultando em 54% de recall e 88% de precisão. O desempenho inferior nessa classe específica pode ser atribuído à ausência de uma gramática especificamente construída para ela, destacando a necessidade de adaptação da abordagem de acordo com as características únicas de cada conjunto de dados.

Entidades nomeadas identificadas automaticamente em textos têm sido usadas em várias aplicações com o objetivo de reduzir o esforço humano em tarefas manuais e repetitivas. Pirovani, Spalenza e Oliveira (2017), por exemplo, usaram 10 tipos de entidades nomeadas identificadas automaticamente usando CRF+LG em textos educacionais para gerar perguntas, economizando o tempo do professor. Eles encontraram 7195 entidades nomeadas e geraram 6917 perguntas (4175 de preenchimento de lacunas e 2742 de texto livre).

Pirovani, Nogueira e Oliveira (2018) também usaram CRF+LG para encontrar nomes de pessoas em um grande conjunto de dados de um jornal. Os nomes encontrados foram usados para criar uma página web de índice de nomes de pessoas. Essa página permite que o usuário acesse facilmente todos os artigos do jornal onde seus nomes aparecem dirija-se diretamente ao ponto onde o nome é citado na página do jornal.

Além disso, entidades nomeadas identificadas em textos também são usadas para aumentar



a compreensão do assunto do texto e melhorar a qualidade de sistemas com diferentes propósitos. Oliveira et al. (2019) usaram entidades nomeadas como características adicionais para aprimorar a qualidade dos sistemas de avaliação automatizada de redações. Com as características de cada redação respectiva, incluindo as entidades nomeadas, as notas são então inseridas como a variável critério em uma regressão linear. O estudo mostrou a importância de levar em consideração, para avaliar a qualidade dos argumentos dos alunos, as entidades nomeadas em relação ao tema abordado na redação. Eles obtiveram uma melhoria em duas das cinco dimensões dos critérios de avaliação usados pelo exame ENEM em relação aos resultados de ponta.

Entidades nomeadas químicas foram usadas por Izo et al. (2023) para gerar relatórios inteligentes para a tomada de decisões a partir de textos livres. Eles utilizaram patentes químicas e artigos científicos, extraíram as entidades nomeadas através do método híbrido CRF+LG e, em seguida, aplicaram regras para gerar informações inteligentes. Foram gerados 35 relatórios inteligentes combinando informações de documentos que citam métodos, equipamentos, compostos, elementos e classes químicas similares.

3 PROPOSTA

Um documento carrega em si muitos elementos textuais que são valiosas informações para alguém, ou alguéms em algum dia. Não podemos prever com certeza, a priori, qual palavra ou conjunto de estruturas textuais complexas, entre muitos outros fenômenos linguísticos textuais, serão de uso valioso para um pesquisador, ou um cidadão, ou mesmo para arquivistas. Por essa razão, precisamos de melhores ferramentas e metodologias para reformular rapidamente nossa estrutura de descrição arquivística a fim de atender às necessidades particulares dos usuários do momento.

Moraes, Zafalon e Barroso (2019) destacam que a descrição arquivística consiste em identificar e elucidar o contexto e o conteúdo dos documentos arquivísticos. Essa atividade é fundamental para alinhar as concepções teóricas e práticas da Ciência da Informação e da Arquivologia, dado que ambas compartilham a informação como objeto de estudo. Nesse sentido, a Tecnologia da Informação e Comunicação desempenha um papel fundamental, permitindo que aplicações de software, como o AtoM, consolidem e interpretem a informação arquivística armazenada. Com isso, este trabalho propõe o uso de inteligência artificial para acelerar o processo de descrição arquivística, indo além dos elementos textuais



tradicionais, como número do processo, nomes das partes e tipo de assunto. A abordagem sugerida examina o documento como um todo, identificando entidades nomeadas, o que aprimora a eficiência e a precisão no acesso dos usuários aos documentos.

Novos algoritmos de inteligência artificial estão à nossa disposição para anotar estruturas linguísticas ricas a partir de textos (Pirovani; Oliveira, 2021). Com base em uma abordagem semelhante ao artigo citado anteriormente, focamos, no presente trabalho, em identificar e anotar alguns elementos de uma coleção de decisões judiciais. Esses elementos comporão os metadados desses documentos ao registrá-los no AtoM. Com esse objetivo, aprimoramos o algoritmo anterior para identificar o tempo da sentença na decisão judicial, entre outros elementos discutidos nos resultados, na Seção 4. Primeiramente, anotamos manualmente 30 decisões para treinar o algoritmo. Durante esse processo manual, marcamos todos os elementos que queremos que o algoritmo aprenda: 1) o número do processo, o documento; 2) a data de emissão da decisão; 3) a comarca; 4) o tribunal; 5) o nome do juiz responsável; 6) o nome do réu, ou réus; e 7) a sentença definitiva. Esta é o tempo da sentença após alguma fração de tempo adicional ou de redução com base nas circunstâncias em que o crime foi cometido. A Figura 1 mostra um exemplo de parte de uma decisão anotada. O réu foi condenado a 6 anos, 8 meses, e uma multa de 16 dias de acordo com 1/30 do maior salário mínimo vigente.



Figura 1: Um exemplo de decisão judicial anotada.

Ante o exposto, JULGO PROCEDENTE o pedido do **Ministério Público** para CONDENAR o réu **TALVANES MAGALHÃES DA SILVA JÚNIOR**, qualificado nos autos, como incurso no artigo **157**, § **2º-A**, inciso I do Código Penal, à pena de **06 (seis)** anos e **08 (oito)** meses de reclusão e **16 (dezesesseis)** dias-multa, calculados à razão de **1/30** do maior salário-mínimo vigente à época dos fatos, com regime semiaberto para início do cumprimento.

Não concedo ao réu o direito de recorrer em liberdade pelos mesmos fundamentos explicitados para a decretação de sua prisão preventiva, por conta da gravidade do crime a ele imputado, para assegurar a ordem pública e garantir a aplicação da lei penal, sobretudo agora diante da condenação, considerando o montante e regime de pena impostos.

Recomende-se o réu ao estabelecimento prisional onde se encontra. Após, com a intimação das partes acerca da presente sentença, expeça-se, com urgência, guia de recolhimento provisória/definitiva.

Por fim, DETERMINO a restituição dos objetos apreendidos à fl. **20** em favor do sentenciado, expedindo-se o necessário após o trânsito em julgado.

Custas conforme a lei (artigo **4º**, § **9º**, "a", da Lei Estadual nº **11.608/2003**), observando-se a gratuidade processual concedida nos autos em favor do sentenciado (fl. **117**).

P. R. e I.

Araçatuba, 15 de setembro de 2021.

Fonte: De autoria própria

Após ter um conjunto de documentos anotados manualmente, os algoritmos propostos utilizaram esses documentos anotados para aprender a anotar automaticamente os documentos que ainda não foram anotados por especialistas.

4 EXPERIMENTOS E RESULTADOS

A anotação de texto continua sendo um desafio para as comunidades científicas, em parte devido ao alto custo de treinar algoritmos para anotar grandes volumes de documentos arquivísticos. Esses algoritmos geralmente exigem um extenso conjunto de dados previamente rotulados (Li et al., 2022), o que demanda um grande esforço de anotação humana e validação cruzada. Nossa proposta oferece uma abordagem mais econômica para o treinamento do algoritmo, que é capaz de aprender rapidamente e oferece desempenho de qualidade estatisticamente comparável aos algoritmos de referência na literatura.

Para capturar dados de treinamento, desenvolvemos um aplicativo de software baseado no

Moodle², que permite a especialistas humanos inserirem suas anotações em dados que serão utilizados no treinamento. O sistema solicita a um grupo selecionado de usuários que anotem uma amostra de documentos, utilizando uma linguagem baseada em XML (*Extensible Markup Language*), um padrão que define um conjunto de regras para codificar documentos em um formato que seja legível tanto por humanos quanto por máquinas, para marcar entidades nomeadas presentes nos documentos (como exemplificado na Figura 2). Esse processo continua até que amostras suficientes sejam acumuladas para treinar o algoritmo.

Figura 2: Um exemplo de decisão judicial anotada.

```

1 Ante o exposto, JULGO PROCEDENTE o pedido do
2 <EM ID="1500164-90.2021.8.26.0603-12" CATEG="ORGANIZACAO">Ministério Público</EM> para CONDENAR o réu
3 <EM ID="1500164-90.2021.8.26.0603-13" CATEG="PESSOA">TALVANES MAGALHÃES DA SILVA JÚNIOR</EM>,
4 qualificado nos autos, como incurso no artigo
5 <EM ID="1500164-90.2021.8.26.0603-14" CATEG="VALOR">157</EM>, §
6 <EM ID="1500164-90.2021.8.26.0603-15" CATEG="VALOR">2º</EM>-A, inciso I do Código Penal, à pena de
7 <EM ID="1500164-90.2021.8.26.0603-16" CATEG="VALOR">06</EM> (
8 <EM ID="1500164-90.2021.8.26.0603-17" CATEG="VALOR">seis</EM>) anos e
9 <EM ID="1500164-90.2021.8.26.0603-18" CATEG="VALOR">08</EM> (
10 <EM ID="1500164-90.2021.8.26.0603-19" CATEG="VALOR">oito</EM>) meses de reclusão e
11 <EM ID="1500164-90.2021.8.26.0603-20" CATEG="VALOR">16</EM> (
12 <EM ID="1500164-90.2021.8.26.0603-21" CATEG="VALOR">dezesesseis</EM>) dias-multa, calculados à razão de
13 <EM ID="1500164-90.2021.8.26.0603-22" CATEG="VALOR">1/30</EM> do maior salário-mínimo vigente à época dos fatos, com
14 regime semiaberto para início do cumprimento. Não concedo ao réu o direito de recorrer em liberdade pelos mesmos fundamentos
15 explicitados para a decretação de sua prisão preventiva, por conta da gravidade do crime a ele imputado, para assegurar a
16 ordem pública e garantir a aplicação da lei penal, sobretudo agora diante da condenação, considerando o montante e regime
17 de pena impostos. Recomende-se o réu ao estabelecimento prisional onde se encontra. Após, com a intimação das partes acerca
18 da presente sentença, expeça-se, com urgência, guia de recolhimento provisória/definitiva.
19 Por fim, DETERMINO a restituição dos objetos apreendidos à fl.
20 <EM ID="1500164-90.2021.8.26.0603-23" CATEG="VALOR">20</EM> em favor do sentenciado, expedindo-se o necessário após
21 o trânsito em julgado. Custas conforme a lei (artigo
22 <EM ID="1500164-90.2021.8.26.0603-24" CATEG="VALOR">4º</EM>, §
23 <EM ID="1500164-90.2021.8.26.0603-25" CATEG="VALOR">9º</EM>, "a", da Lei Estadual nº
24 <EM ID="1500164-90.2021.8.26.0603-26" CATEG="VALOR">11.608/2003</EM>), observando-se a gratuidade processual
25 concedida nos autos em favor do sentenciado (fl.
26 <EM ID="1500164-90.2021.8.26.0603-27" CATEG="VALOR">117</EM>). P. R. e I.
27 <EM ID="1500164-90.2021.8.26.0603-28" CATEG="LOCAL">Araçatuba</EM>,
28 <EM ID="1500164-90.2021.8.26.0603-29" CATEG="TEMPO">15 de setembro de 2021</EM>.
29 DOCUMENTO ASSINADO DIGITALMENTE NOS TERMOS DA LEI
30 <EM ID="1500164-90.2021.8.26.0603-30" CATEG="VALOR">11.419/2006</EM>, CONFORME IMPRESSÃO À MARGEM DIREITA
31 <EM ID="1500164-90.2021.8.26.0603-31" CATEG="PESSOA">Kathleen Freeman</EM>

```

Fonte: De autoria própria

Em termos de resultados, o algoritmo apresentou uma precisão de 0,7 e uma medida F de 0,49. Esses valores indicam que, embora o algoritmo seja capaz de identificar corretamente uma boa parte das entidades nomeadas (precisão), ele ainda apresenta uma necessidade de melhoria na harmonia entre precisão e recall (medida F). Isso sugere que, embora o modelo tenha uma boa capacidade de evitar falsos positivos, ainda há espaço para otimizar o equilíbrio entre a precisão e a capacidade de detectar todas as entidades relevantes. No entanto, considerando que o treinamento foi realizado com um número limitado de amostras, esses resultados são promissores e indicam que o algoritmo tem potencial para melhorar com

² <http://moodle.org/>



o aumento do conjunto de dados de treinamento.

5 CONSIDERAÇÕES FINAIS

Propõe-se neste trabalho uma abordagem com o uso de inteligência artificial para acelerar a tarefa de descrição arquivística. Mais do que apenas prestar atenção aos elementos textuais tradicionais, como por exemplo o número do processo, os nomes das partes e o tipo do assunto em disputa, nossa abordagem busca no documento inteiro por entidades nomeadas para melhorar a capacidade de acesso dos usuários aos seus documentos.

Aumentar a qualidade dos metadados arquivísticos com novas e mais complexas expressões textuais, como entidades nomeadas, permite melhorias na busca em sistemas de informação, como já discutido pelo autores Izo, Oliveira e Badue (2021). Portanto, o usuário é amplamente beneficiado pela incorporação dessa tecnologia em um sistema de informação. No trabalho atual, estudamos os marcos iniciais para fornecer essas funcionalidades aos documentos de decisões judiciais. Com a incorporação dessa funcionalidade pode-se inclusive inferir-se a classificação dos documentos baseando-se apenas nos valores das sentenças.

Um trabalho futuro que planejamos realizar é a extração de palavras, ou expressões, que expliquem o aumento ou diminuições das sentenças pelos juízes. Ou seja, busca-se assim as razões e explicações dos motivos que sustentaram a atribuição das sentenças. Por outro lado, mecanismos como os que buscamos jogará luzes a que novas metodologias possam ser estudadas para estabelecer diretrizes e critérios claros para os juízes seguirem ao determinar suas sentenças.

REFERÊNCIAS

Campos, J.; Oliveira, E. **Extração de Nomes de Pessoas em Textos em Português: uma Abordagem Usando Gramáticas Locais**. In: Computer on the Beach 2015. Florianópolis, SC: SBC, 2015. Best paper award. Disponível em: <[https:// hal.archives-ouvertes.fr/ hal-01134971](https://hal.archives-ouvertes.fr/hal-01134971)>.

Colombo, C.; Oliveira, E. **Intelligent Information System for Extracting Knowledge from Pharmaceutical Package Inserts**. In: XVIII Brazilian Symposium on Information Systems (SBSI). Florianópolis, SC: SBC, 2022.



Conselho Nacional de Arquivos (CONARQ). **Resolução nº 51, de 11 de julho de 2023.** Estabelece diretrizes para a implementação de repositórios arquivísticos digitais confiáveis (RDC-Arq). Disponível em: <https://www.gov.br/conarq/pt-br/resolucoes/resolucao-51.pdf>. Acesso em: 08 ago. 2024.

Izo, F.; Oliveira, E.; Badue, C. **Named Entities as a Metadata Resource for Indexing and Searching Information.** In: 21th International Conference on Intelligent Systems Design and Applications – (ISDA). On the WWW: Springer International Publishing, 2021. v. 418, p. 838–848.

Izo, F. et al. **An Intelligent Report Generator for Chemical Documents.** In: XIX Brazilian Symposium on Information Systems. Curitiba, Brazil: Association for Computing Machinery, 2023. p. 276–283.

Li, Q. et al. **A Survey on Text Classification: From Traditional to Deep Learning.** ACM Transactions on Intelligent Systems and Technology (TIST), ACM New York, NY, v. 13, n. 2, p. 1–41, 2022.

Manning, C. D.; Schuetze, H. **Foundations of Statistical Natural Language Processing.** 1st. ed. New York, NY: The MIT Press, 1999.

Moraes, H. A. R.; Zafalon, Z. R.; Barroso, T. DE B. **Descrição arquivística, Records in Contexts (RiC) e Access to Memory (AtoM): análise exploratória da literatura científica.** RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação, v. 17, n. 0, p. e019009, 24 jan. 2019.

Oliveira, E. et al. **The Influence of NER on the Essay Grading.** In: SPRINGER. 19th International Conference on Intelligent Systems Design and Applications: Intelligent Systems Design and Applications. Delhi, India: Springer International Publishing, 2019. p. 102–113.

Pirovani, J.; Nogueira, M.; Oliveira, E. **Indexing Names of Persons in a Newspaper Large**



Dataset. In: 13th International Conference on the Computational Processing of Portuguese (PROPOR). Canela, RS: Springer, 2018. v. 11122.

Pirovani, J.; Oliveira, E. **Studying the Adaptation of Portuguese NER for Different Textual Genres.** The Journal of Supercomputing, Springer International Publishing, p. 1–17, 2021.

Pirovani, J.; Spalenza, M.; Oliveira, E. **Geração Automática de Questões a Partir do Reconhecimento de Entidades Nomeadas em Texto Didáticos.** In: XXVIII Simpósio Brasileiro de Informática na Educação (SBIE). Ceará, CE: SBC, 2017. p. 1147–1156.

Santos, C. N.; Guimaraes, V. **Boosting Named Entity Recognition with Neural Character Embeddings.** In: Proceedings of the Fifth Named Entities Workshop, ACL 2015. Stroudsburg, PA, USA: Association for Computational Linguistics, 2015. p. 25–33.

Sutton, C.; McCallum, A. **Conditional Random Fields: An Introduction.** Foundations and Trends® in Machine Learning, Elsevier, v. 4, p. 267–373, 2012. ISSN 1935-8237.