



# **CENTRO UNIVERSITÁRIO LUTERANO DE PALMAS**

*Recredenciado pela Portaria Ministerial nº 1.162, de 13/10/16, D.O.U. nº 198, de 14/10/2016*  
**AELBRA EDUCAÇÃO SUPERIOR - GRADUAÇÃO E PÓS-GRADUAÇÃO S.A.**

Vinícius Dias dos Santos

FERRAMENTA PARA ANOTAÇÃO DE ENTIDADES NOMEADAS

Palmas – TO

2021

Vinicius Dias dos Santos  
FERRAMENTA PARA ANOTAÇÃO DE ENTIDADES NOMEADAS

Trabalho de Conclusão de Curso (TCC) II elaborado e apresentado como requisito parcial para obtenção do título de bacharel em Sistemas de Informação pelo Centro Universitário Luterano de Palmas (CEULP/ULBRA).

Orientador: Prof. Esp. Fábio Castro Araújo.

Palmas – TO

2021

Vinícius Dias dos Santos

FERRAMENTA PARA ANOTAÇÃO DE ENTIDADES NOMEADAS

Trabalho de Conclusão de Curso (TCC) II elaborado e apresentado como requisito parcial para obtenção do título de bacharel em Sistemas de Informação pelo Centro Universitário Luterano de Palmas (CEULP/ULBRA).

Orientador: Prof. Esp. Fábio Castro Araújo.

Aprovado em: \_\_\_\_/\_\_\_\_/\_\_\_\_

BANCA EXAMINADORA

---

Prof. Esp. Fábio Castro Araújo

Orientador

Centro Universitário Luterano de Palmas – CEULP

---

Prof. Me. Jackson Gomes de Souza

Centro Universitário Luterano de Palmas – CEULP

---

Profª. A Dra. Parcilene Fernandes de Brito

Centro Universitário Luterano de Palmas – CEULP

Palmas – TO

2021

## **AGRADECIMENTOS**

Gostaria de agradecer e dedicar este trabalho as seguintes pessoas:

Primeiramente a Deus que permitiu que tudo isso acontecesse, ao longo de minha vida, e não somente nestes anos como universitário.

À minha mãe Delza Dias dos Santos e minha segunda mãe Joana Dias dos Santos pelo apoio incondicional aos meus estudos.

Ao meu orientador Fábio Castro, pelos grandes ensinamentos ao longo deste trabalho e curso, bem como a relação de amizade que foi estabelecida.

Agradeço aos meus amigos, em especial os meus antigos colegas, pelos ótimos momentos que vivemos ao longo deste curso, que nos transformou em uma equipe verdadeiramente completa.

A todos que direta ou indiretamente fizeram parte de minha formação, o meu muito obrigado.

## RESUMO

SANTOS, Vinícius Dias dos. **FERRAMENTA PARA ANOTAÇÃO DE ENTIDADES NOMEADAS**. 2021. 37 f. Trabalho de Conclusão de Curso (Graduação) - Curso de Sistemas de Informação, Computação, Centro Universitário Luterano de Palmas, Palmas, 2021.

Anotar entidades consiste em identificar objetos do mundo real presentes em textos, no qual o usuário seleciona palavras e dá a elas uma categoria. Existem categorias que são consideradas genéricas, devido a sua facilidade de identificação, como locais, pessoas, tempo, etc. Este trabalho apresenta a construção de uma ferramenta *web* para a tarefa de anotação e classificação de entidades nomeadas em textos. No referencial teórico são apresentados alguns conceitos relacionados a esta tarefa, como o de extração de dados e reconhecimento de entidades nomeadas em textos. Na construção da ferramenta estão presentes o *framework* Angular e as bibliotecas Angular Material e PrimeNG. A metodologia seguida neste trabalho é dividida em algumas etapas, sendo elas a estruturação da anotação e classificação, bem como a estruturação do arquivo JSON a ser exportado no formato sugerido pelo PubAnnotation. Como resultados deste trabalho estão os três módulos desenvolvidos. O primeiro módulo é o de barra de navegação, que tem a responsabilidade de controlar a navegação e também gerar o arquivo de exportação de entidades. O segundo módulo é o de categorias, que gerencia as categorias que o usuário utiliza para classificar entidades. O terceiro e último se trata do módulo de anotação, sendo este o responsável por disponibilizar as funcionalidades para que o usuário anote entidades no texto. A ferramenta desenvolvida cumpre o papel de auxiliar os usuários iniciantes no momento de gerar os dados das anotações.

Palavras-chave: Entidades Nomeadas, Anotação.

## LISTA DE FIGURAS

|   |    |
|---|----|
| Figura 1 - Anotação e classificação de entidades                              | 13 |
| Figura 2 - Estrutura básica de entidades - Modelo PubAnnotation               | 14 |
| Figura 3 - Sequência de Métodos Utilizados.                                   | 17 |
| Figura 4 - Esboço da Seleção de Entidades.                                    | 17 |
| Figura 5 - Estrutura das Entidades Exportadas em JSON.                        | 19 |
| Figura 6 - Estrutura dos Componentes.   | 20 |
| Figura 7 - Barra de Navegação   | 21 |
| Figura 8 - Opções de Navegação  | 21 |
| Figura 9 - Comunicação entre Componentes                                      | 22 |
| Figura 10 - Arquivo de Exportação Gerado Pela Ferramenta                      | 22 |
| Figura 11 - Componente de gerenciamento de categorias                         | 23 |
| Figura 12 - Criação de uma nova categoria                                     | 24 |
| Figura 13 - Componente de anotação  | 25 |
| Figura 14 - Menu de contexto com categorias                                   | 26 |
| Figura 15 - Representação do evento de seleção de texto                       | 27 |
| Figura 16 - Representação dos índices após anotação.                          | 27 |
| Figura 17 - Conteúdo da área de texto após a marcação de uma entidade         | 28 |
| Figura 18 - Obtenção do elemento anterior através do evento de seleção        | 29 |
| Figura 19 - Fluxograma da obtenção dos índices reais da entidade selecionada. | 30 |
| Figura 20 - Alteração dos índices das entidades após anotação.                | 31 |
| Figura 21 - Obtenção dos índices da entidade anotada com base no innerHTML    | 32 |
| Figura 22 - Manipulação do DOM para anotar entidades                          | 32 |

## **LISTA DE TABELAS**

Tabela 1 - Características das ferramentas de anotação existentes.

**10**

## **LISTA DE ABREVIATURAS E SIGLAS**

PLN – Processamento de Linguagem Natural

REN – Reconhecimento de Entidades Nomeadas

EI – Extração de Informação

JSON – *JavaScript Object Notation*

DOM – *Document Object Model*

## SUMÁRIO

|           |  |           |
|-----------|--|-----------|
| <b>1.</b> | <b>INTRODUÇÃO</b>                            | <b>8</b>  |
| <b>2.</b> | <b>REFERENCIAL TEÓRICO</b>                   | <b>12</b> |
| 2.1.      | EXTRAÇÃO DE DADOS DA WEB                     | 12        |
| 2.2.      | RECONHECIMENTO DE ENTIDADES NOMEADAS         | 12        |
| <b>3.</b> | <b>METODOLOGIA</b>                           | <b>16</b> |
| <b>4.</b> | <b>RESULTADOS E DISCUSSÃO</b>                | <b>20</b> |
| 4.1.      | ARQUITETURA DA FERRAMENTA                    | 20        |
| 4.1.1.    | BARRA DE NAVEGAÇÃO E EXPORTAÇÃO DE ENTIDADES | 21        |
| 4.1.2.    | CATEGORIAS                                   | 23        |
| 4.2.      | ANOTAÇÃO                                     | 24        |
| <b>5.</b> | <b>CONSIDERAÇÕES FINAIS</b>                  | <b>34</b> |
|           | <b>REFERÊNCIAS</b>                           | <b>35</b> |

## 1. INTRODUÇÃO

Conforme as tecnologias de informação avançaram, a quantidade de dados textuais produzidos pela humanidade aumentou, principalmente devido ao advento da internet. Grande parte deles não são estruturados, ou seja, organizados com uma estrutura previamente planejada, o que dificulta seu entendimento por máquinas (PRIVATTO, 2020). Tendo em vista a necessidade de manipular estes dados para a extração de informações relevantes tem-se a técnica de Reconhecimento de Entidades Nomeadas (REN) que é uma sub tarefa dentro da tarefa de Extração de Informações (EI) (DIAS et al., 2020).

Essa sub tarefa possibilita a identificação de trechos importantes em textos, mesmo que estes não estejam estruturados, através da busca e classificação de palavras-chave, sendo as entidades mais comuns: pessoas, locais, tempo e organizações (DIAS et al., 2020). Dentro do REN existem diversos métodos para anotar essas entidades presentes nos textos, alguns fazem uso de técnicas de aprendizado profundo para identificação de entidades de uma área específica, com base em textos já anotados. Em sua dissertação, Oliveira (2020) faz uso deste método para reconhecer entidades em narrativas de roubos. Com base em um conjunto rotulado de textos, a rede neural implementada em seu trabalho é treinada a fim de obter o máximo possível de entidades importantes para o domínio abordado, ao qual obteve sucesso na identificação de 36 tipos de entidades, automatizando assim o processo de obtenção de informações sobre crimes violentos contra o patrimônio.

Outras alternativas são ferramentas de anotação manual, que consiste em ler um documento e informar onde está cada palavra-chave para que possa ser feita a classificação da mesma em seguida. A anotação manual é considerada um empecilho no Processamento de Linguagem Natural (PLN), por ser geralmente um processo demorado, porém deve-se levar em consideração a ferramenta utilizada para tal, visto que a qualidade da mesma interfere diretamente na produtividade de quem vai anotar (NEVES e SEVA, 2021).

Devido a isso, diversas ferramentas com este propósito são regularmente lançadas e, muitas vezes, para se escolher qual utilizar é imprescindível testá-las antes para entender suas funcionalidades. Este passo pode requerer um certo esforço devido à dificuldade de instalação, encontrar código fonte ou devido a aplicação necessitar de uma assinatura para uso (NEVES e SEVA, 2021).

Este trabalho visa implementar uma ferramenta para anotação e classificação de entidades nomeadas, requerendo apenas o uso de um navegador para sua utilização, para tentar entregar ao usuário um ambiente produtivo. A ferramenta abordada se trata de um *software* que permitirá que os usuários anotem e classifiquem entidades em um texto de sua escolha. A anotação ocorre de forma manual através de uma interface que reconhece quando uma entidade foi selecionada pelo usuário, permitindo também a atribuição de uma categoria para a entidade selecionada.

Com base nisso, este trabalho visa responder o seguinte problema de pesquisa: como implementar uma ferramenta que possibilite a anotação e classificação de entidades nomeadas em textos?

Analisando o problema de pesquisa, a hipótese que este trabalho considera para resolvê-lo seria através da utilização de tecnologias para desenvolvimento *web*, como o Angular. Com o auxílio dessas tecnologias *web* é possível criar uma interface na qual o usuário possa inserir um determinado texto e que também permite que o usuário selecione entidades para anotar e classificar. Sendo assim, este trabalho visa desenvolver um *software* que disponibiliza ao usuário funcionalidades que o permite anotar entidades nomeadas por meio da seleção de palavras em textos.

A cada ano, diversos *softwares* que possuem funcionalidades para reconhecimento de entidades nomeadas são implementados, sendo grande parte deles disponibilizados por meio de um navegador, como mostra a Tabela 1, que exhibe alguns dos *softwares* analisados por Neves e Seva (2021) que obtiveram maior nota na análise feita e que ainda se encontram disponíveis para utilização.

**Tabela 1 - Características das ferramentas de anotação existentes.**

| <b>Ferramenta</b>    | <b>Características</b>  | <b>Disponibilidade</b>   | <b>Gratuito</b> |
|----------------------|---|--|-----------------|
| WebAnno <sup>1</sup> | <ul style="list-style-type: none"> <li>● Código aberto;</li> <li>● Importa textos em diferentes formatos de arquivos;</li> <li>● Exporta anotações em diversos tipos de formatos;</li> <li>● Suporte à criação de projetos e verificação de anotação por um curador.</li> </ul> | <ul style="list-style-type: none"> <li>● Requer instalação;</li> <li>● Baseado na <i>web</i>.</li> </ul> | Sim             |

<sup>1</sup> <https://webanno.github.io/webanno/>

|                        |   |  |                     |
|------------------------|---|--|---------------------|
| Prodigy <sup>2</sup>   | <ul style="list-style-type: none"> <li>● Fácil instalação;</li> <li>● Suporte a vários formatos de entrada de texto, porém com uma prévia estruturação dos mesmos.</li> <li>● Exporta anotações em JSON;</li> <li>● Faz pré-anotação de entidades utilizando aprendizado de máquina.</li> </ul>   | <ul style="list-style-type: none"> <li>● Requer instalação;</li> <li>● Baseado na <i>web</i>.</li> </ul> | Não                 |
| BioQRator <sup>3</sup> | <ul style="list-style-type: none"> <li>● Fácil instalação;</li> <li>● Importa arquivos no formato BioC e PubMed;</li> <li>● Suporta exportação de entidades anotadas no formato BioC e CSV;</li> <li>● Permite importação de nomenclaturas;</li> <li>● Os dados são armazenados nos servidores da ferramenta;</li> <li>● Não tem suporte a vários idiomas.</li> </ul>   | <ul style="list-style-type: none"> <li>● Disponível online.</li> </ul>                                   | Sim                 |
| Doccano <sup>4</sup>   | <ul style="list-style-type: none"> <li>● Possui outras funcionalidades além da anotação, como análise de sentimentos e classificação de imagens;</li> <li>● Importa arquivos no formato JSON, texto plano ou CoNLL, o mesmo vale para a exportação;</li> <li>● Código aberto;</li> <li>● Suporta anotações em equipes;</li> <li>● Suporte a vários idiomas;</li> </ul>  | <ul style="list-style-type: none"> <li>● Requer instalação;</li> <li>● Baseado na <i>web</i>.</li> </ul> | Sim                 |
| Tagtog <sup>5</sup>    | <ul style="list-style-type: none"> <li>● Possui limitações na quantidade de usuários, tipos de entidades, projetos entre outras coisas em sua versão gratuita;</li> <li>● Importa textos em diferentes formatos de arquivos;</li> <li>● Exporta anotações em diversos formatos;</li> <li>● Permite a sobreposição de anotações (anotar uma parte de uma entidade que já foi anotada);</li> <li>● Suporte a multiusuários e anotação automática em sua versão paga.</li> </ul> | <ul style="list-style-type: none"> <li>● Disponível online (através de <i>login</i>).</li> </ul>         | Sim, com limitações |
| TextAE <sup>6</sup>    | <ul style="list-style-type: none"> <li>● Código aberto;</li> <li>● Importa arquivos JSON no formato PubAnnotation, o mesmo vale para a exportação.</li> <li>● Aceita também um <i>link</i> para uma</li> </ul>  | <ul style="list-style-type: none"> <li>● Disponível online.</li> </ul>                                   | Sim                 |

<sup>2</sup> <https://prodi.gy/>

<sup>3</sup> <http://www.bioqrator.org/>

<sup>4</sup> <http://doccano.herokuapp.com/>

<sup>5</sup> <https://www.tagtog.net/>

<sup>6</sup> <http://textae.pubannotation.org/>

|  |  |  |  |
|--|--|--|--|
|  | anotação presente no repositório<br>PubAnnotation. |  |  |
|--|--|--|--|

Com base nas características apresentadas na Tabela 1 e na análise de Neves e Seva (2021), constata-se que grande parte dos *softwares* possuem poucas ou nenhuma publicação abordando sobre sua construção ou funcionamento. O estudo aponta também que a grande maioria dessas ferramentas não é atualizada há mais de dois anos, sendo que, atualmente, muitas delas foram descontinuadas e não se encontram mais disponíveis para utilização. Verifica-se também o destaque do WebAnno com relação às publicações feitas sobre ele, as funcionalidades que o mesmo disponibiliza, além de outros aspectos técnicos, como importação e exportação de dados.

Devido a sua alta gama de recursos, o WebAnno requer que o usuário faça algumas configurações após a instalação da ferramenta, e necessita também de um estudo prévio da sua documentação. Tudo isso pode resultar em dificuldades para usuários iniciantes que buscam apenas anotar textos, pois para usufruir da ferramenta necessita-se a leitura prévia de sua documentação. Diante deste cenário, é de grande utilidade uma ferramenta atualizada que possibilita que seus usuários anotem seus textos de forma objetiva, para que este não sinta necessidade de recorrer a leitura de uma extensa documentação para realizar anotações simples.

## 2. REFERENCIAL TEÓRICO

### 2.1. EXTRAÇÃO DE DADOS DA WEB

A internet possui quantidade gigantesca de dados que se expande ano após ano devido a facilidade de publicação de informações que seus usuários possuem nela. Essa facilidade, embora vantajosa para o crescimento desses dados, gera também dificuldades para utilização dos mesmos em algum contexto. Visando a busca por informações relevantes na *web*, áreas como a Extração de Dados surgiram. Esta área auxilia na localização, coleta e organização de dados presentes na *web*, estruturando-os em um formato de saída conveniente para uso por outras aplicações (PONTOLIO, 2014).

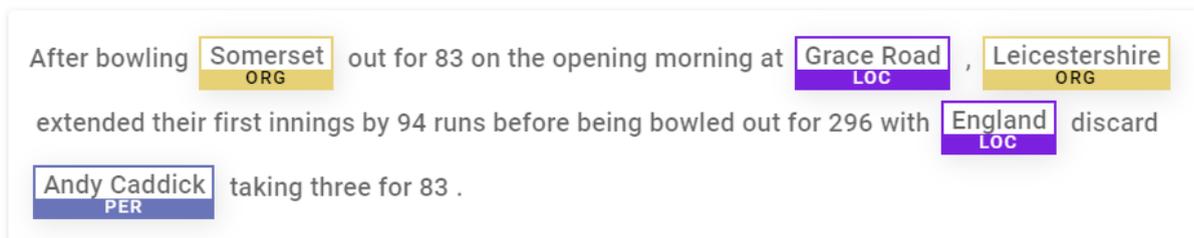
Dentre as técnicas de extração de dados, é comum o uso de robôs ou *scrapers* (como são conhecidos) que simulam a interação humana em determinados sites. A vantagem dessa técnica é a possibilidade de extrair grandes quantidades de dados desses sites, sendo superior a uma pessoa fazendo o mesmo procedimento. Esta técnica possui uma alta dependência com o layout de cada site, pois um *scraper* é ensinado a manipular um certo *layout* para nele obter as informações, e, caso haja uma mudança no mesmo, pode acarretar em problemas durante essa extração. Embora seja uma dependência arriscada, a utilização de *scrapers* ainda é uma alternativa viável em diversos casos, se tratando de extração de informações da web, principalmente devido a sua velocidade e precisão (SILVA e BARION, 2019).

### 2.2. RECONHECIMENTO DE ENTIDADES NOMEADAS

O Reconhecimento de Entidades Nomeadas (REN) é uma tarefa dentro da subárea de Extração de Informação (EI), cujo principal propósito é o de identificar e classificar entidades em textos, desde os tipos mais genéricos de entidades como Pessoa, Organização, Local, Tempo, até os mais específicos, que variam com o contexto a qual a tarefa é aplicada (AMARAL; VIEIRA, 2014). A Figura 1 demonstra como entidades podem ser anotadas e classificadas em um determinado texto através de uma demarcação, indicando onde inicia e termina a entidade, também a rotulando com uma categoria, como os exemplos mostrados

“Grace Road” rotulada como local, “Andy Caddick” como pessoa e “Somerset” como organização.

**Figura 1 - Anotação e classificação de entidades**



Fonte: Doccano, 2021.

Para esta tarefa existem diversas abordagens, sendo as principais as baseadas em regras, de aprendizado de máquina e também as que utilizam em conjunto ambas as citadas (AMARAL; VIEIRA, 2014). O REN através de algoritmos de aprendizado de máquina consiste na criação de modelos probabilísticos que identificam padrões em grandes volumes de dados. O aprendizado desses modelos podem ser ou não supervisionados. Em aprendizado não supervisionado os algoritmos trabalham com critérios, analisando os dados com base neles. No aprendizado supervisionado os dados contêm rótulos, ou seja, são exemplos de como deveria ser a resposta, sendo assim os algoritmos buscam correlacionar as análises com os rótulos contidos nesses dados, permitindo a classificação automática de novos casos parecidos (PINTO, 2015).

Embora o foco na abordagem por aprendizado de máquina tenha aumentado, a necessidade de dados anotados manualmente por humanos para identificação de entidades nomeadas se torna uma desvantagem, pois esses dados servem de base para o treinamento desses algoritmos. Já as abordagens baseadas em regras não necessitam desses dados anotados, pois fazem essa identificação com base no conhecimento do especialista que dita as regras ao algoritmo, o que leva à uma relação de dependência do mesmo, para que sempre as mantenha atualizadas. Apesar de ferramentas que fazem uso de aprendizado de máquina terem um desempenho superior, ainda é necessário ressaltar que estas soluções geralmente são para contextos extremamente específicos e sem planejamento para possíveis expansões para outros domínios, devido à alta complexidade (FERRI, 2016).

Estes dados anotados, nesta área de extração de informações, são comumente chamados de *corpus*, que se trata de um conjunto de textos com entidades já identificadas

(SANTANA, 2020). Os *corpus* anotados são divididos em duas categorias, existindo assim o *corpus* dourado e prateado, sendo o primeiro um conjunto de entidades anotadas ou revisadas por um ser humano, o que torna sua construção mais difícil, e por consequência fazendo com que esse tipo de corpora se torne mais escasso que o prateado. O prateado são as anotações decorrentes de algum tipo de técnica computacional, como aprendizado de máquina, regras, etc, que, normalmente, faz uso de um *corpus* dourado para seu treinamento (SILVA, 2020).

Para auxiliar na construção de mais *corpus* dourados, ferramentas para anotação manual têm surgido, como a TALEN, que se trata de um *software web* que tem como foco principal auxiliar o anotador no momento de anotar textos de línguas das quais o mesmo não domina ou não possui conhecimento. O *software* foi um dos pioneiros no ramo de ferramentas de anotação para não falantes do idioma do texto a ser anotado. Em comparação com a ferramenta de anotação Brat, os anotadores classificaram o TALEN como mais eficiente após obterem resultados satisfatórios (MAYHEW e ROTH, 2018).

Analisando as ferramentas para anotação existentes, listadas por Neves e Seva (2021), observa-se que o formato JSON é comumente utilizado como forma de exportação das entidades que foram anotadas e classificadas. O JSON tem um formato leve, no qual os valores contidos nele são precedidos por uma chave que o nomeia (BOURHIS et al., 2017). Embora as informações contidas nos arquivos exportados não sigam exatamente a mesma estrutura, elas ainda contêm os valores presentes no formato definido pelo PubAnnotation. O PubAnnotation se trata de um repositório de código aberto que possui uma API e também um site que permite que qualquer usuário possa criar e enviar anotações de texto.

**Figura 2 - Estrutura básica de entidades - Modelo PubAnnotation**

```
{
  "text": "IRF-4 expression in CML may be induced by IFN-α therapy",
  "denotations": [
    {"id": "T1", "span": {"begin": 0, "end": 5}, "obj": "Protein"},
    {"id": "T2", "span": {"begin": 42, "end": 47}, "obj": "Protein"}
  ]
}
```

Fonte: PubAnnotation, 2021.

A Figura 2 exibe o formato seguido pelas anotações submetidas ao repositório PubAnnotation, que organiza seu conteúdo em chaves-valores, conforme a sintaxe do JSON. A estrutura conta com uma chave denominada “*text*” que tem o propósito de armazenar o texto original antes da anotação. Observa-se também a existência de uma chave “*denotations*” que contém uma lista de entidades que foram anotadas. Essa lista pode conter vários objetos, sendo cada um deles a representação de uma entidade anotada. Cada um desses objetos contém uma identificação, chave “*id*”, além de um objeto “*span*”, que contém as chaves “*begin*” e “*end*” que armazenam, respectivamente, o índice inicial e final da entidade anotada. Além das chaves mencionadas, o objeto da anotação mostrado contém também uma última chave chamada “*obj*”, cujo propósito é guardar a categoria da entidade, que foi definida durante a classificação da mesma.

### 3. METODOLOGIA

Para desenvolver este trabalho, foram utilizadas algumas das tecnologias disponíveis na área de desenvolvimento *web*, sendo elas: o *framework* Angular, o mecanismo de estilização *Sassy Cascading Style Sheets* (SCSS) e as bibliotecas de componentes Angular Material e PrimeNG, que juntas compõem o *front-end* da aplicação, ou seja, a interface que o usuário visualiza e interage.

O Angular é uma plataforma de desenvolvimento criada e mantida pelo Google que oferece um conjunto de bibliotecas e ferramentas integradas para criação de aplicações *web* escaláveis. O *framework* oferece nativamente componentes documentados que auxiliam a implementar grande parte dos *layouts* das telas, como botões, listas, menus, etc, além de permitir que componentes de terceiros sejam adicionados ao projeto.

PrimeNG e Angular Material se tratam de bibliotecas de componentes de interface do usuário para o *framework* Angular. As bibliotecas fornecem desde os componentes mais básicos como botões, caixas de seleção, entradas de texto até mais complexos, como caixas de seleção com funcionalidade de autocompletar textos, gráficos, *captchas*, etc. O diferencial em utilizar estas bibliotecas está na estrutura dos componentes disponibilizados por elas, que permitem que sejam personalizados de acordo com as necessidades de quem o utiliza.

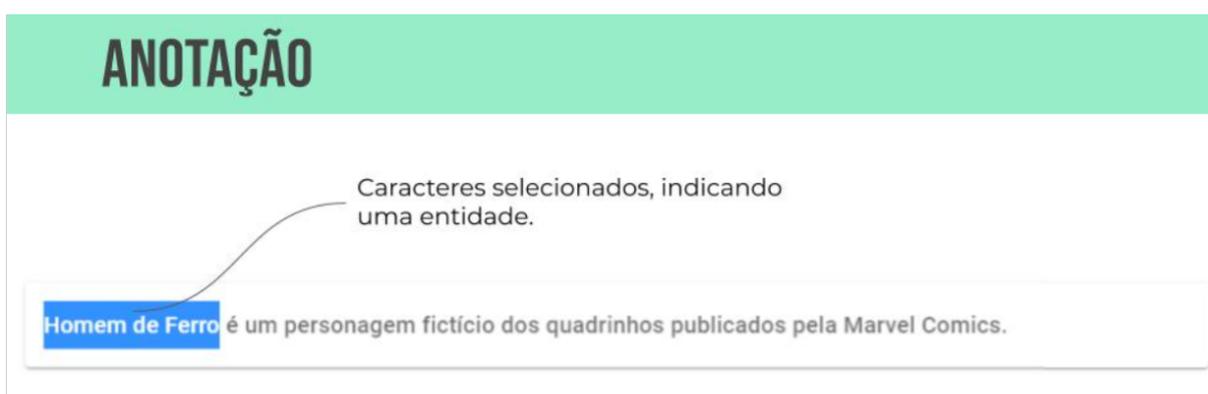
Para auxiliar no desenvolvimento da ferramenta, alguns métodos foram elaborados, conforme mostra a Figura 3. A Figura mostra que o desenvolvimento da ferramenta foi realizado em etapas, implementando uma funcionalidade de cada vez.

Figura 3 - Sequência de Métodos Utilizados.



A primeira etapa consiste em definir como será possível o usuário anotar e classificar entidades a partir do texto importado. A ferramenta possibilitará que o usuário execute esta ação através da seleção de caracteres, com isso ele pode definir que uma ou mais palavras possam ser anotadas como uma só entidade, conforme ilustrado na Figura 4.

Figura 4 - Esboço da Seleção de Entidades.



Após fazer a seleção dos caracteres, a ferramenta exibirá um menu contendo categorias genéricas e também as que o usuário definiu anteriormente. Essas categorias servem para classificar a entidade selecionada. Este menu permite também que o usuário crie uma nova categoria a qualquer momento caso assim deseje.

Além da anotação e classificação, foi preciso também elaborar uma maneira de exportar as entidades anotadas, o que foi feito na segunda etapa conforme mostrado na Figura 3. Essa exportação visa facilitar a manipulação dos textos anotados por outras aplicações/algoritmos de aprendizado de máquina. Nos dados suplementares<sup>7</sup> da publicação de Neves e Seva (2021), constata-se que o JSON é o formato mais utilizado para exportação dentre as ferramentas de anotação existentes. Visando uma maior compatibilidade, a ferramenta implementada neste trabalho utiliza este formato para exportação das entidades anotadas. A estrutura dos dados anotados é similar ao modelo definido pelo PubAnnotation, que faz uso do JSON, sendo assim os dados são organizados através da chave/valor. Nestes dados estão contidos o texto sem anotações e também uma lista das anotações realizadas nele. Este modelo é amplamente utilizado por ferramentas de anotação como Docanno e também bibliotecas de PLN como a Spacy, que utiliza esses dados como treinamento de modelos existentes. O arquivo JSON exportado contém informações do texto fonte e informações das entidades anotadas, informando seu índice inicial e final no texto, bem como sua categoria, decorrente da classificação do usuário, conforme mostra a Figura 5. O atributo “*text*” representa o texto original, que o usuário importará na ferramenta para anotá-lo. Já o atributo “*labels*” guarda as informações do que foi anotado. Um texto pode conter vários rótulos, sendo assim este atributo é uma lista, contendo a localização da entidade no texto e também sua categoria.

---

<sup>7</sup> NEVES, Mariana; SEVA, Jurica. Supplementary material for the publication: An extensive review of tools for manual annotation of documents. **Briefings In Bioinformatics**. Oxford, p. 146-163. jan. 2021. Disponível em: <https://bit.ly/3C1KGu5>. Acesso em: 24 abr. 2021.

**Figura 5 - Estrutura das Entidades Exportadas em JSON.**

```
[
  {
    "text": "Almira é uma ótima pessoa, gosto muito dela",
    "labels": [
      [0, 6, "PER"]
    ]
  },
  {
    "text": "O casamento, que era dia 12/06/2020, foi adiado para o dia 20/07/2020",
    "labels": [
      [25, 35, "DATE"],
      [59, 68, "DATE"]
    ]
  },
  {
    "text": "O sonho dela era ir para Austrália visitar seu irmão",
    "labels": [
      [25, 34, "LOC"]
    ]
  }
]
```

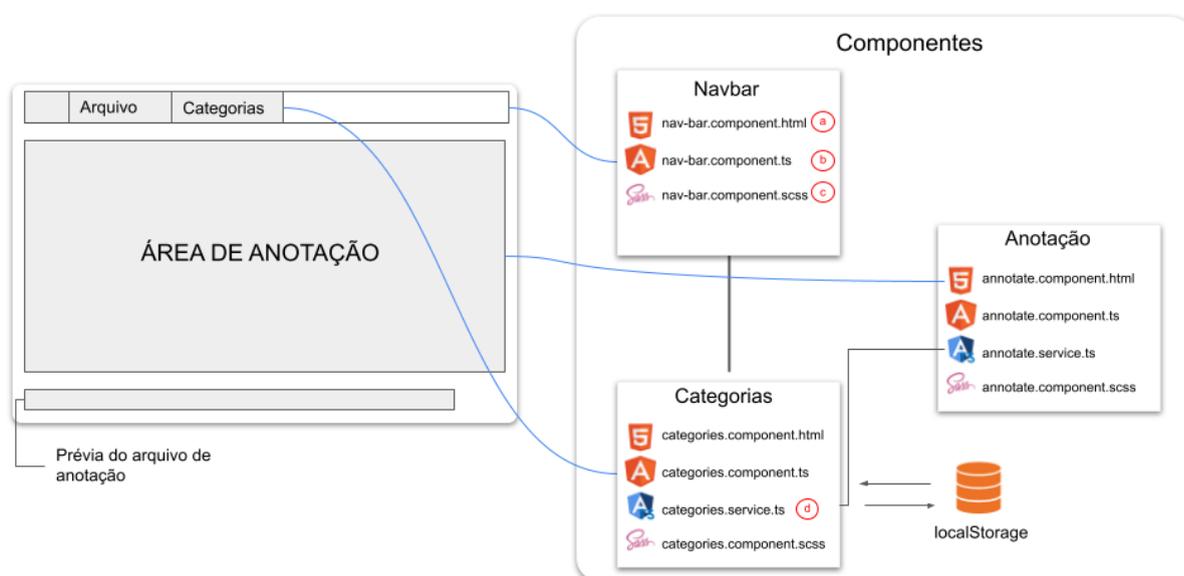
Devido a maioria das funcionalidades se concentrarem na interface que é exibida ao usuário, grande parte do trabalho é feito no Angular, através da criação de componentes para exibição do texto importado, seleção de caracteres, exibição de categorias para classificação de entidades. O SCSS será utilizado para estilizar as telas da ferramenta, incluindo a anotação de entidades, tendo a responsabilidade de atribuir cores diferentes para entidades de categorias distintas, com o intuito de melhorar a experiência do usuário ao anotar entidades. Durante a última etapa exibida na Figura 3 todas as funcionalidades descritas nas etapas anteriores serão mescladas em uma ferramenta que permitirá a anotação de textos e a exportação das entidades anotadas.

## 4. RESULTADOS E DISCUSSÃO

### 4.1. ARQUITETURA DA FERRAMENTA

Por se tratar de um projeto no *framework* Angular, a estrutura da ferramenta consiste em diferentes componentes independentes que se comunicam através de serviços para troca de informações. Conforme mostra a Figura 6, cada componente possui uma estrutura similar, contando cada um com quatro arquivos.

Figura 6 - Estrutura dos Componentes.

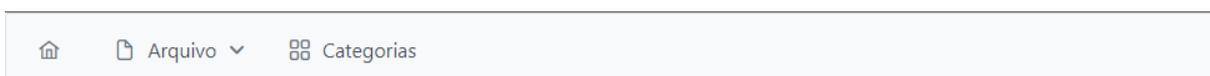


O primeiro arquivo, representado na figura pela letra “a”, se trata do responsável pela exibição do conteúdo ao usuário final. O segundo, representado pela letra “b”, gerencia toda a lógica e controle do comportamento do componente de exibição do primeiro arquivo. O terceiro arquivo, simbolizado na letra “c”, é o que realiza a obtenção de dados necessários para o funcionamento do componente e também o responsável por manter a lógica de negócios isolada do segundo arquivo. Por fim, no quarto arquivo, simbolizado pela letra “d”, é onde foi implementada toda a estilização dos elementos exibidos pelo primeiro arquivo, tais como tamanhos de fonte, cores e etc.

#### 4.1.1. BARRA DE NAVEGAÇÃO E EXPORTAÇÃO DE ENTIDADES

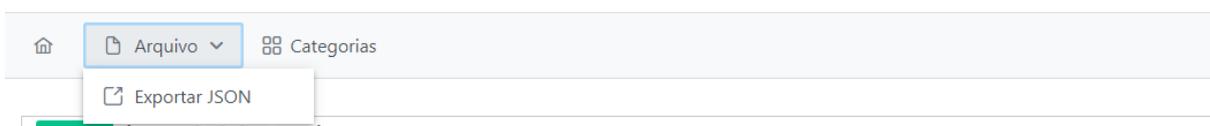
Para manter a organização das funcionalidades presentes na ferramenta, foi criado um componente cuja finalidade é a exibição de uma barra de navegação. Este elemento, conforme ilustra a Figura 7, exibe opções para que o usuário faça a exibição de outros componentes ou execute determinadas ações.

**Figura 7 - Barra de Navegação**



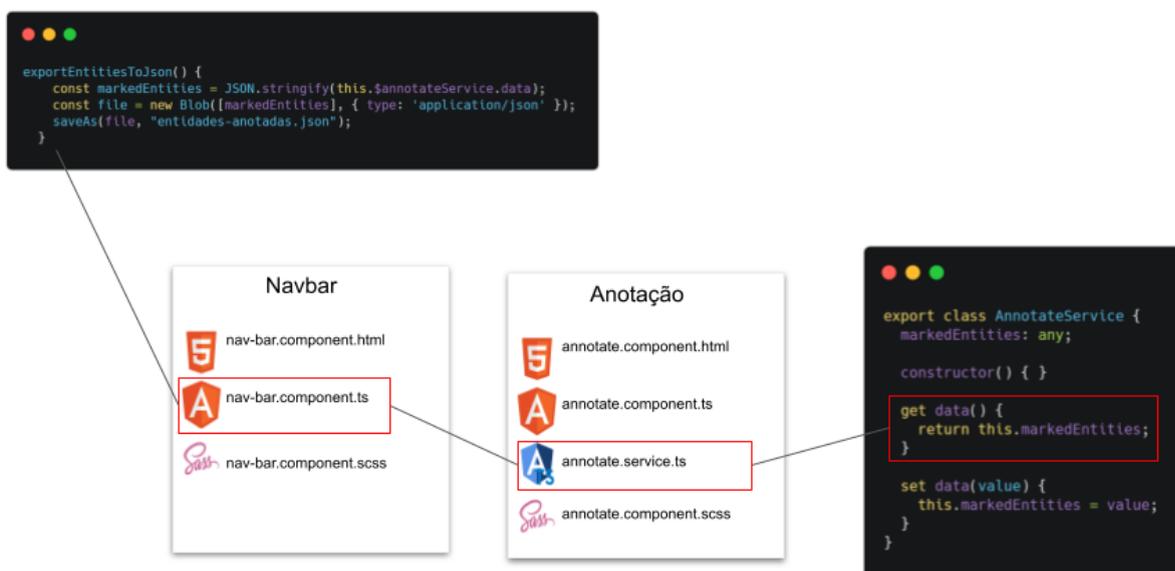
Dentre as ações mencionadas, está a opção “Exportar JSON” exibida na Figura 8, que, ao ser pressionada, aciona uma funcionalidade presente no componente que obtém as entidades anotadas no componente de anotação e faz o *download* de um arquivo JSON com as entidades obtidas.

**Figura 8 - Opções de Navegação**



Essa comunicação entre este componente e o componente de anotação, exemplificada pela Figura 9, acontece por meio da injeção de dependências. A função “*exportEntitiesToJson*”, presente no componente `NavBar`, faz uso de um serviço implementado no componente de anotação, que está preparado para disponibilizar todas as palavras que foram anotadas no texto. Com isso, a função consegue obter todos os dados necessários para construir e efetuar o *download* do arquivo contendo as entidades que foram anotadas já no formato JSON.

Figura 9 - Comunicação entre Componentes



A Figura 10 mostra como o arquivo se apresenta após ser aberto. O arquivo mantém a estrutura proposta durante a metodologia deste trabalho apresentada na Figura 5.

Figura 10 - Arquivo de Exportação Gerado Pela Ferramenta

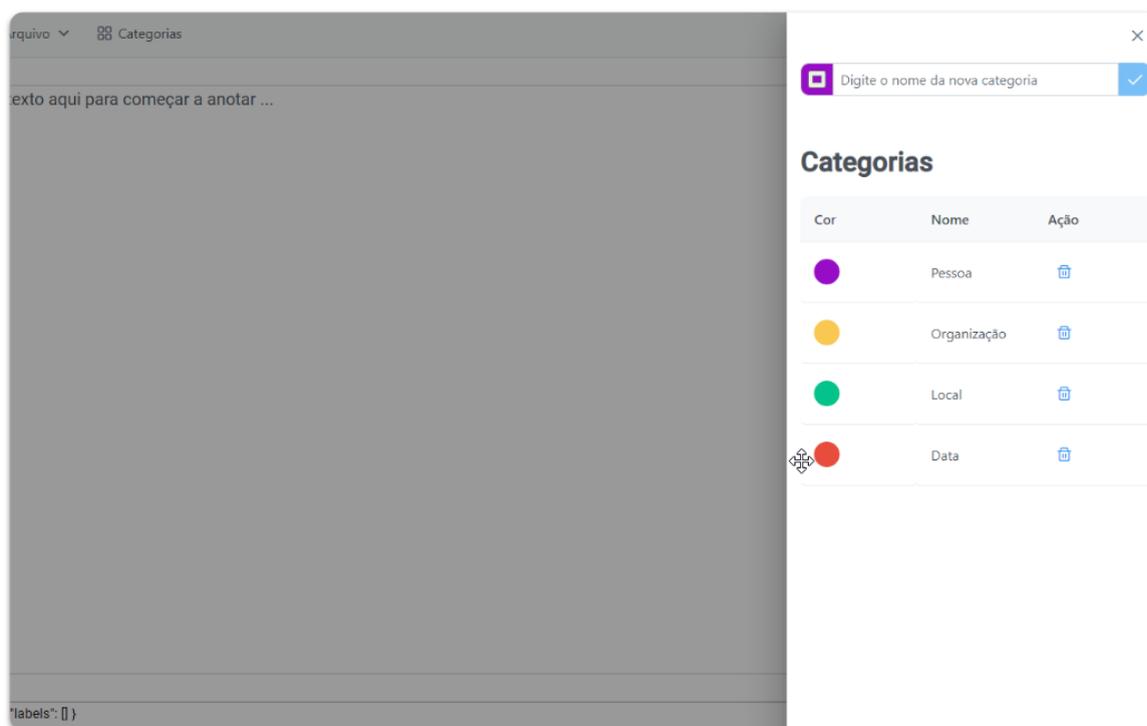
```
1 {
2   "sourceText": "Fortaleza é a capital do Ceará",
3   "labels": [
4     [0, 9, "Local"],
5     [25, 30, "Local"]
6   ]
7 }
```

Essa estrutura conta um atributo “*sourceText*” que tem o propósito de armazenar o texto fonte fornecido pelo usuário na área de texto. Há também o atributo “*labels*” que armazena todas as entidades anotadas em uma lista, sendo cada item desta lista composto pelo início e fim da entidade e logo em seguida a categoria que lhe foi atribuída.

#### 4.1.2. CATEGORIAS

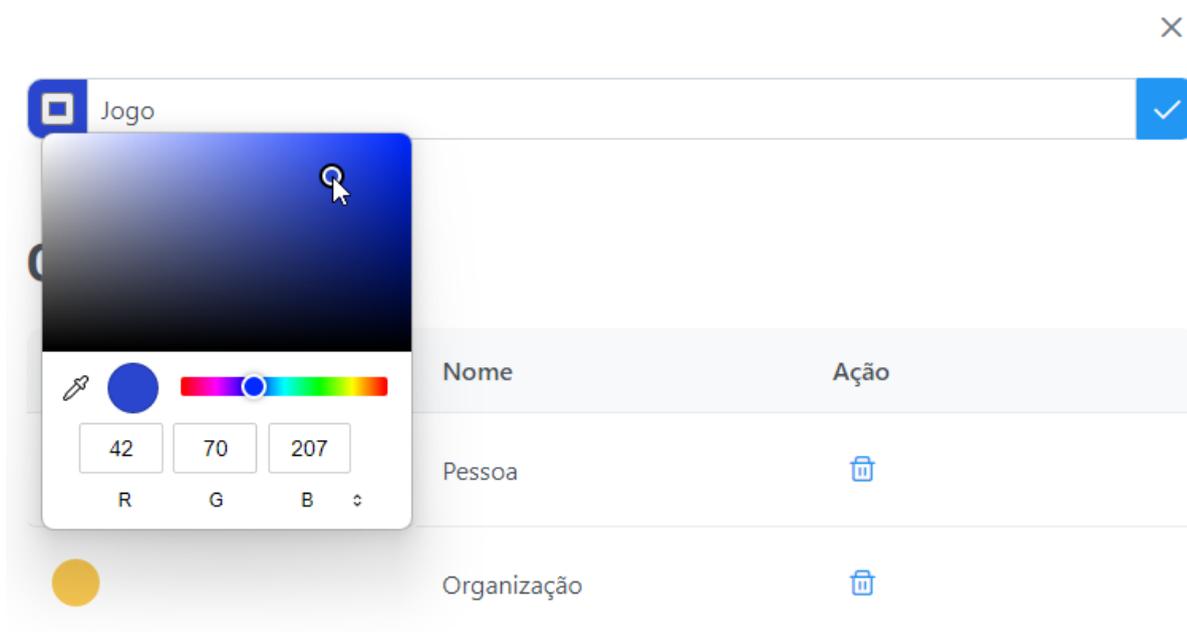
Antes ou durante um procedimento de anotação, o usuário pode definir categorias que serão atribuídas às entidades anotadas. Para isso foi implementado o componente de categorias, cujo objetivo é gerenciar a exibição e também o armazenamento dessas categorias. O componente, ilustrado pela Figura 11, exibe as categorias salvas no *localStorage*, que se trata de um armazenamento local destinado ao navegador. Por padrão, o componente criará automaticamente as entidades genéricas “Pessoa”, “Organização”, “Local” e “Data”, visando permitir que o usuário realize uma anotação sem nunca antes ter definido uma categoria.

**Figura 11 - Componente de gerenciamento de categorias**



Conforme mostrado na Figura 12, o componente exibe também uma opção para criar novas categorias. Para realizar essa criação, é necessário que o usuário informe um nome no campo correspondente e também selecione uma cor para a mesma.

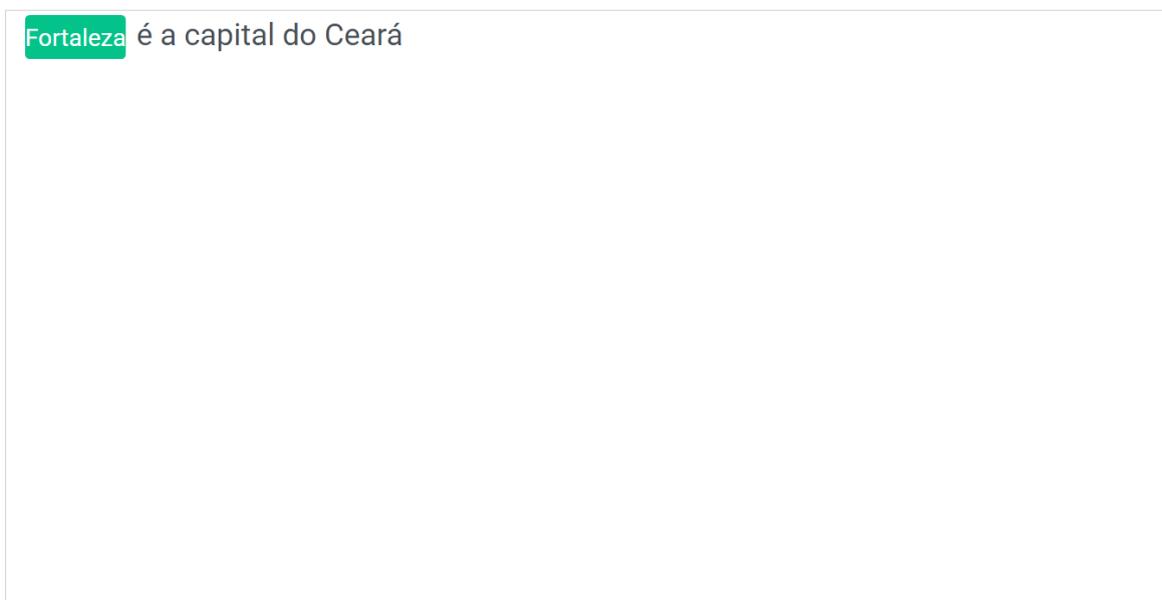
Figura 12 - Criação de uma nova categoria



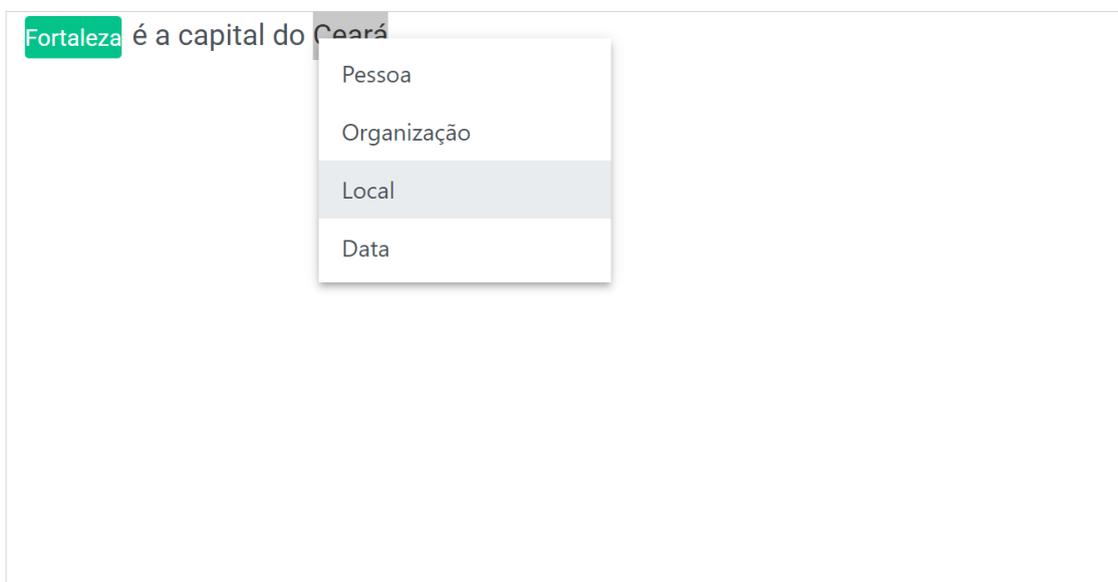
Esta cor será utilizada na marcação da entidade no texto, o que proporciona ao usuário realizar a identificação do que se trata cada marcação. Caso o usuário tente criar uma categoria com o mesmo nome de uma existente a ferramenta emitirá um alerta, visto que cada categoria deve ser única.

#### 4.2. ANOTAÇÃO

Para tornar possível a anotação do texto pelo usuário, foi criado um componente de anotação, conforme mostrado na Figura 13. A área de texto presente neste componente não se trata de um elemento *textarea* e sim de um elemento “*div*” com a propriedade “*contentEditable=true*” aplicada, o que faz com que este elemento se torne editável, semelhante a um *textarea*. A escolha de se utilizar esse tipo de *container* se dá devido ao *textarea* a renderização de *tags* HTML em seu conteúdo, o que torna inviável exibir as marcações corretamente com esta abordagem. Na parte inferior do componente é exibida também uma pré-visualização da anotação realizada, para que o usuário consiga ter uma noção de como ficará o arquivo que será exportado.

**Figura 13 - Componente de anotação**

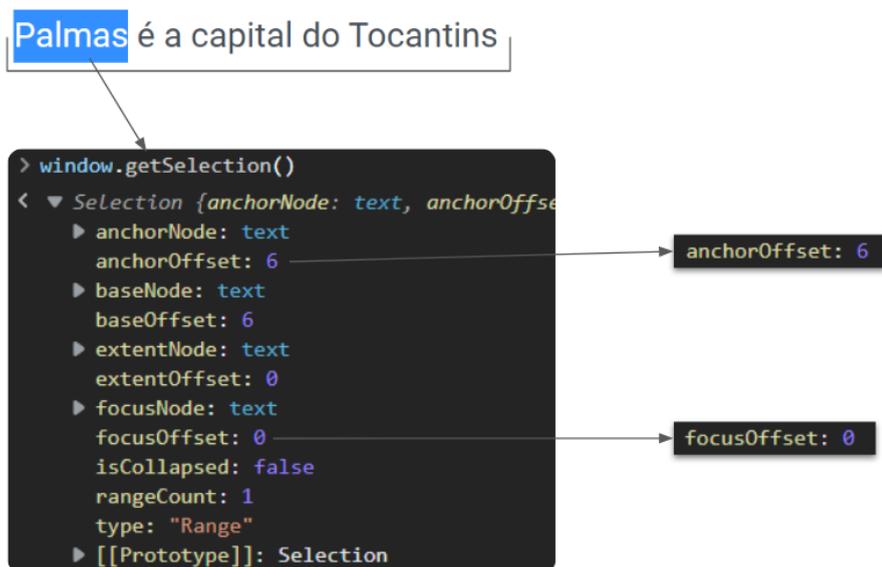
A área de texto permite que o usuário cole ou digite um texto e, em seguida, selecione caracteres inseridos, processo esse ilustrado pela Figura 14. Quando o usuário efetua a seleção, um menu de contexto é exibido para que seja atribuída uma categoria a entidade selecionada. Após o usuário fazer a escolha da categoria, a ferramenta irá realizar o processo de anotação. O processo de anotação requer a obtenção do índice inicial e final da palavra selecionada de acordo com o texto fonte fornecido.

**Figura 14 - Menu de contexto com categorias**

```
{ "sourceText": "Fortaleza é a capital do Ceará", "labels": [[ 0, 9, "Local" ]]}
```

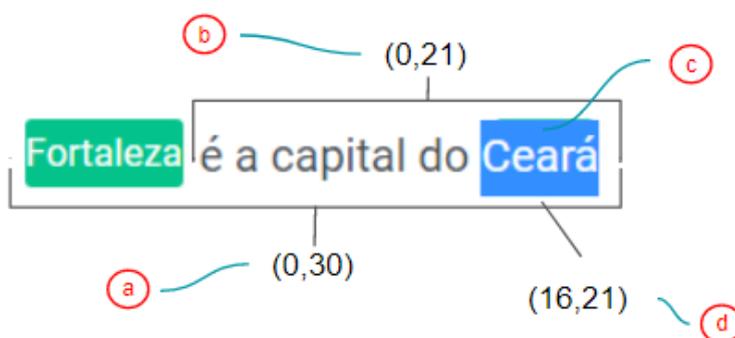
A obtenção do índice original da palavra no texto se dá através de uma funcionalidade que trabalha com o evento de seleção do navegador. A Figura 15 exhibe um objeto JSON que é retornado pela obtenção deste evento de seleção. Este objeto conta com informações detalhadas a respeito da seleção realizada no texto, destacando-se os atributos “*anchorOffset*”, que representa o índice em que a seleção se iniciou, e “*focusOffset*”, sendo este o índice em que a seleção finalizou. Com isso, é possível determinar a posição inicial e final da entidade anotada.

Figura 15 - Representação do evento de seleção de texto



Embora o evento de seleção contenha os índices que marcam o início e fim da entidade selecionada, foi necessário implementar uma nova abordagem de obtenção desses índices. Essa necessidade ocorreu devido a modificação do conteúdo presente na área de texto após realizar uma anotação, como representado pela Figura 16. Na figura é possível observar que o texto completo “Fortaleza é a capital do Ceará” vai do índice “0” (inicial) ao índice “30” (final), conforme exibido em “a”. Entretanto, o evento de seleção da entidade “Ceará” retorna os índices 16 e 21 ilustrado em “d”, sendo que os índices reais dessa entidade seriam 25 e 30.

Figura 16 - Representação dos índices após anotação.



Essa alteração nos índices ocorre devido a anotação realizada anteriormente estar separada do elemento em que a seleção de texto foi efetuada em “c”, fazendo com que o elemento de texto resultante após a marcação possua o índice inicial 0 e final 21, exibidos em “b”. Essa separação de elementos é causada pela adição da *tag* de marcação ao se anotar uma entidade, conforme mostrado na Figura 17.

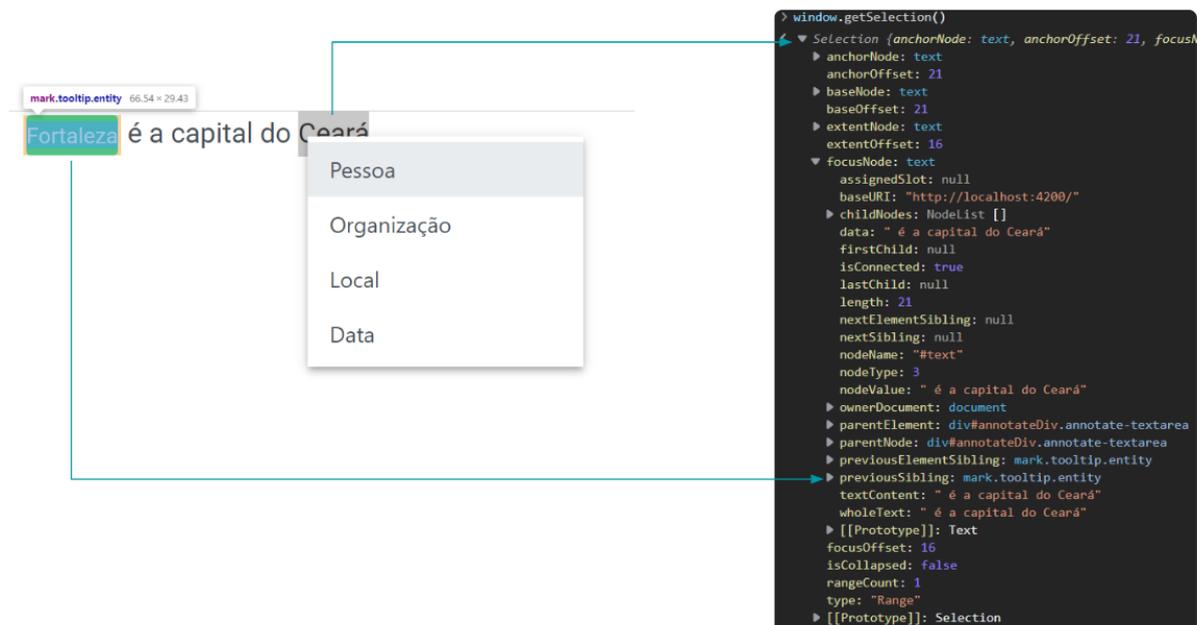
Para o evento de seleção, após a marcação, não existe mais apenas um elemento de texto, e sim dois elementos, um representando a marcação da entidade e outro o texto restante que ainda não foi marcado, ilustrado na Figura 17. Sendo assim os índices do evento são baseados no elemento em que houve a seleção, e não no conteúdo como um todo.

**Figura 17 - Conteúdo da área de texto após a marcação de uma entidade**



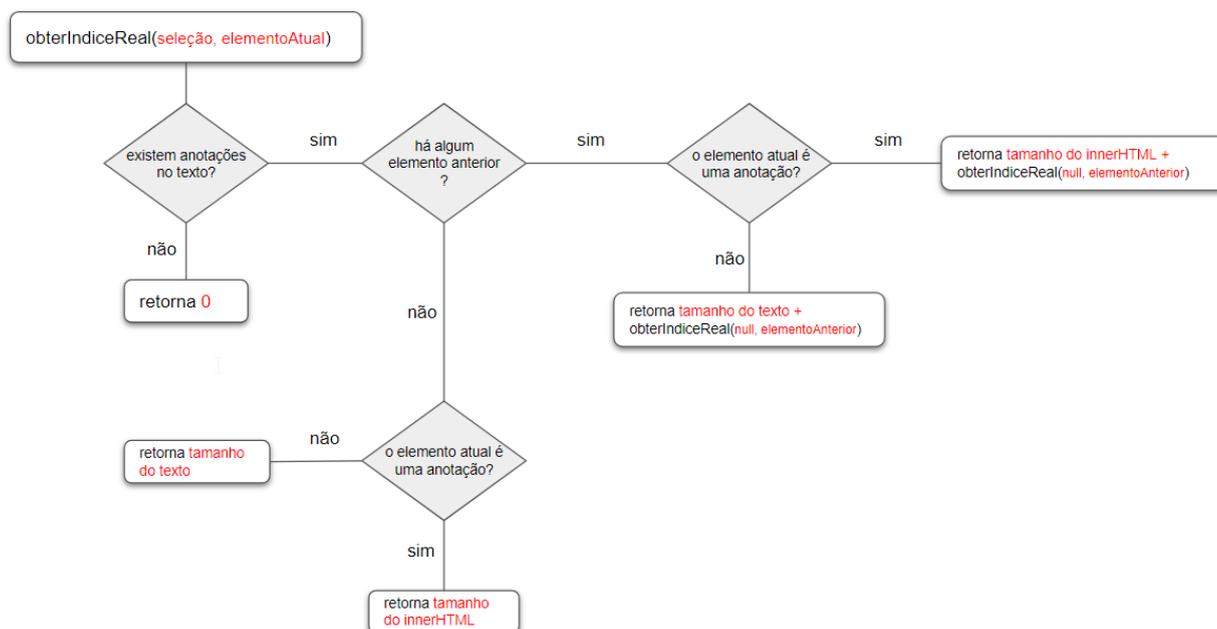
Para contornar o problema, foi implementada uma função que trabalha com o objeto de seleção. Este objeto possui o atributo “*previousSibling*”, que armazena a referência do elemento anterior a ele, conforme mostra a Figura 18. Com isso, foi implementada uma abordagem recursiva sobre este evento, para obter o tamanho dos elementos ancestrais para assim determinar a verdadeira posição inicial e final do texto selecionado.

Figura 18 - Obtenção do elemento anterior através do evento de seleção



Esta função, representada pelo fluxograma da Figura 19, possui dois parâmetros, sendo o primeiro o objeto do evento de seleção de texto e o segundo um objeto auxiliar, que representa elementos anteriores ao elemento em que houve a seleção de texto. A função, através da recursividade, navega até o último elemento anterior ao da seleção e soma o tamanho dos textos de cada elemento durante a navegação.

**Figura 19 - Fluxograma da obtenção dos índices reais da entidade selecionada.**



Após a obtenção do tamanho retornado pela função exibida na Figura 19, este valor é somado ao valor dos índices obtidos pelo evento de seleção. Com isso, é possível determinar qual será o índice inicial da seleção levando em consideração o texto como um todo, e não somente ao elemento selecionado.

Para realizar a marcação da entidade no texto o componente identifica a localização dos índices do texto referente a entidade dentro do atributo *innerHTML*, que se trata de uma propriedade do DOM (*Document Object Model*) que possui todo o texto e também todo o HTML presente em um elemento. Essa identificação é necessária devido aos índices dentro do *innerHTML* sofrerem uma alteração após a manipulação do mesmo para adição das *tags* de anotação, conforme observa-se na Figura 20.

**Figura 20 - Alteração dos índices das entidades após anotação.**

## ÍNDICES ANTES DA ANOTAÇÃO

Fortaleza é a capital do Ceará

(25,30)

## ÍNDICES APÓS A ANOTAÇÃO

(0,146)

Fortaleza é a capital do Ceará

(162,167)

Essa identificação, mostrada na Figura 21, leva como base os índices da entidade selecionada obtidos anteriormente e os soma com o tamanho ocupado pelas entidades cujo índice final é anterior ao da entidade selecionada. O condicional, exibido na linha 5, verifica se há alguma entidade já anotada, caso não exista não será necessária nenhuma modificação nos índices. Caso exista, será necessário verificar se essas entidades marcadas são anteriores ao texto selecionado, verificação essa feita na linha 7. Para saber com precisão o tamanho das entidades anteriores, foi necessário obter também a categoria que foi atribuída a ela, devido ao nome da categoria e também a sua cor fazerem parte do elemento, o que torna possível sua estilização. Com isso, o tamanho da identificação da cor e nome da categoria são somados a um valor fixo “125”, que representa o tamanho das demais *tags* presentes no elemento de marcação, como exibido na linha 15. Com esta soma obtém-se então o índice inicial da marcação. O índice final é calculado na linha 20, ao qual soma-se o índice inicial com o tamanho do texto selecionado. Após a execução desta função a ferramenta consegue então realizar com precisão a substituição da palavra selecionada por uma *tag* de marcação.

Figura 21 - Obtenção dos índices da entidade anotada com base no *innerHTML*

```

1  setIndexesWithAnnotatedText() {
2    this.entityMarkInitialIndex = this.entityInitialIndex;
3    this.entityMarkFinalIndex = this.entityFinalIndex;
4
5    if (this.markedEntities.labels.length > 0) {
6      for (let entity of this.markedEntities.labels) {
7        if (entity[1] < this.entityFinalIndex) {
8          let categoryNameLength = 0;
9          let categoryColorLength = 0;
10         let category = this.$categoriesService.getCategoryByName(entity[2]);
11         if (category) {
12           categoryNameLength = category.name.length;
13           categoryColorLength = category.color.length;
14         }
15         this.entityMarkInitialIndex += 125 + categoryNameLength + categoryColorLength;
16       }
17     }
18
19     this.entityMarkFinalIndex = (
20       this.entityMarkInitialIndex + (this.entityFinalIndex - this.entityInitialIndex)
21     );
22   }
23 }

```

Com todos os índices obtidos, o componente realiza então a marcação da entidade através da alteração do conteúdo presente no *innerHTML* da área de texto, como mostra a Figura 22.

Figura 22 - Manipulação do DOM para anotar entidades.

```

1  markEntity(category: any) {
2    if (this.selectedText) {
3      let innerHTML = this.annotatedTextDiv.nativeElement.innerHTML;
4
5      this.annotatedTextDiv.nativeElement.innerHTML = (
6        innerHTML.substring(0, this.entityMarkInitialIndex) +
7        `` +
9        innerHTML.substring(this.entityMarkFinalIndex)
10     );
11
12     this.addEntityToList(category.name);
13     this.hideContextMenu(this.contextMenu);
14   }
15
16   this.entityMarkInitialIndex = 0;
17   this.entityMarkFinalIndex = 0;
18 }

```

No trecho entre as linhas 5 e 9 pode-se ver a forma com que a entidade anotada é envolvida pela *tag* de marcação e como, após realizada a marcação, a mesma é adicionada à lista de entidades marcadas, processo esse presente na linha 11.

## 5. CONSIDERAÇÕES FINAIS

O propósito deste trabalho foi desenvolver uma aplicação *web* que possibilite aos usuários anotar entidades nomeadas em um texto. O que motivou a criação desta ferramenta foi o fato de que as principais aplicações similares possuem um custo ou têm uma alta curva de aprendizado para sua utilização.

Para desenvolver a ferramenta proposta foi utilizado o *framework* Angular juntamente com as bibliotecas de componentes Angular Material e PrimeNG. Com o auxílio dos componentes disponibilizados por essas bibliotecas, foram desenvolvidos novos componentes responsivos que permitem ao usuário inserir um determinado texto, selecionar caracteres e classificá-los com determinadas categorias.

O maior tempo do desenvolvimento foi destinado a anotação de entidades. Isso se deve a dificuldade de se disponibilizar uma área de texto que seja editável e que permita a renderização de *tags* HTML em seu conteúdo. A constante manipulação do DOM exigiu a criação de uma abordagem para uma contínua obtenção de índices das entidades selecionadas. Isso se deve para que as entidades sejam anotadas corretamente e suas posições, tanto a inicial quanto a final, fossem devidamente mantidas conforme o texto que foi inserido na ferramenta.

Visando melhorar a aplicação, tem-se como trabalhos futuros a melhora da abordagem da obtenção dos índices desenvolvida bem como a forma com que as marcações são renderizadas na área de texto. Essa melhora tem o intuito de permitir que o usuário edite uma anotação realizada, o que não é possível na abordagem atual.

Outro trabalho futuro seria a criação de uma API para processar o texto inserido na ferramenta. Esta API faria uso de *machine learning* para realizar esse processamento. Com o uso desta técnica, a aplicação poderia, de forma automática, identificar entidades no texto inserido e sugerir anotações para o usuário, que por sua vez poderia aceitá-las ou recusá-las.

## REFERÊNCIAS

AMARAL, Daniela Oliveira F. do; VIEIRA, Renata. NERP-CRF uma ferramenta para o reconhecimento de entidades nomeadas por meio de Conditional Random Fields.

**Linguamática**, [S. I.], v. 6, n. 1, p. 41-49, jul. 2014. Semestral. Disponível em: [https://www.inf.pucrs.br/linatural/wordpress/wp-content/uploads/2017/10/NERP\\_CRF\\_Linguamatica.pdf](https://www.inf.pucrs.br/linatural/wordpress/wp-content/uploads/2017/10/NERP_CRF_Linguamatica.pdf). Acesso em: 25 abr. 2021.

BOURHIS, Pierre et al. JSON: data model, query languages and schema specification. In: **Proceedings of the 36th ACM SIGMOD-SIGACT-SIGAI symposium on principles of database systems**. 2017. p. 123-135.

DIAS, M. et al. Named Entity Recognition for Sensitive Data Discovery in Portuguese. **Applied Sciences**, v. 10, n. 7, p. 2303, 27 mar. 2020. Disponível em: <https://www.mdpi.com/2076-3417/10/7/2303>. Acesso em: 24 abr. 2021.

FERRI, Junior. **Abordagem modular baseada em dicionário para reconhecimento de entidades nomeadas através de associação aproximada**. 2016. 72 f. Dissertação (Mestrado) - Curso de Ciência da Computação, Universidade Federal do Paraná, Curitiba, 2016. Disponível em: <https://acervodigital.ufpr.br/bitstream/handle/1884/46487/R%20-%20D%20-%20JUNIOR%20FERRI.pdf?sequence=1&isAllowed=y>. Acesso em: 25 abr. 2021.

MAYHEW, Stephen; ROTH, Dan. Talen: Tool for annotation of low-resource entities. In: **Proceedings of ACL 2018, System Demonstrations**. 2018. p. 80-86. Disponível em: <https://www.aclweb.org/anthology/P18-4014.pdf>. Acesso em: 25 abr. 2021.

NEVES, Mariana; SEVA, Jurica. An extensive review of tools for manual annotation of documents. **Briefings In Bioinformatics**. Oxford, p. 146-163. jan. 2021. Disponível em: <https://doi.org/10.1093/bib/bbz130>. Acesso em: 24 abr. 2021.

OLIVEIRA, Bárbara Stéphanie Neves. **APRENDIZADO PROFUNDO PARA RECONHECIMENTO DE ENTIDADES NOMEADAS EM NARRATIVAS DE ROUBOS**. 2020. 99 f. Dissertação (Mestrado) - Curso de Ciência da Computação,

Universidade Federal do Ceará, Quixadá, 2020. Disponível em:

[http://repositorio.ufc.br/bitstream/riufc/55717/1/2020\\_tcc\\_bsnoliveira.pdf](http://repositorio.ufc.br/bitstream/riufc/55717/1/2020_tcc_bsnoliveira.pdf). Acesso em: 30 jul. 2021.

PINTO, Adriano Costa. **CONTROLE INTEGRADO DE TENSÃO E POTÊNCIA REATIVA ATRAVÉS DE APRENDIZADO DE MÁQUINA**. 2015. 150 f. Dissertação (Mestrado) - Curso de Engenharia Elétrica, Universidade Estadual de Campinas, Campinas, 2015. Disponível em:

[http://repositorio.unicamp.br/bitstream/REPOSIP/259987/1/Pinto\\_AdrianoCosta\\_M.pdf](http://repositorio.unicamp.br/bitstream/REPOSIP/259987/1/Pinto_AdrianoCosta_M.pdf). Acesso em: 15 jul. 2021.

PONTOLIO, Luan Silveira. **Plataforma de Extração e Recuperação de Dados na Web no Contexto de Big Data**. 2014. 75 f. TCC (Graduação) - Curso de Sistemas de Informação, Centro Universitário Eurípides de Marília, Marília, 2014. Disponível em:

<https://aberto.univem.edu.br/bitstream/handle/11077/1006/Luan%20Silveira%20Pontolio.pdf?sequence=1&isAllowed=y>. Acesso em: 30 jul. 2021.

PRIVATTO, Pedro Ivo Monteiro. **Uma Abordagem para Reconhecimento de Entidades Nomeadas usando Conhecimento Externo**. 2020. 28 f. Dissertação (Mestrado) - Curso de Ciência da Computação, Universidade Estadual Paulista, Rio Claro, 2020. Disponível em:

<https://repositorio.unesp.br/handle/11449/194224>. Acesso em: 21 abr. 2021.

SANTANA, Jairo Brandão de. **DESENVOLVIMENTO E ANÁLISE DE CORPUS PARA RECONHECIMENTO DE ENTIDADES NOMEADAS EM RELATÓRIOS DE**

**INTELIGÊNCIA FINANCEIRA**. 2020. 104 f. Dissertação (Doutorado) - Curso de Ciência da Informação, Universidade Federal de Santa Catarina, Florianópolis, 2020. Disponível em: <https://repositorio.ufsc.br/bitstream/handle/123456789/219522/PCIN0241-D.pdf?sequence=-1&isAllowed=y>. Acesso em: 14 jul. 2021.

SILVA, Andressa Vieira e. **Um modelo de classificação para o Reconhecimento de Entidades Nomeadas**. 2020. 132 f. Dissertação (Mestrado) - Curso de Letras, Linguística, Universidade de São Paulo, São Paulo, 2020. Disponível em:

[https://www.teses.usp.br/teses/disponiveis/8/8139/tde-06042021-192617/publico/2020\\_AndressaVieiraESilva\\_VCorr.pdf](https://www.teses.usp.br/teses/disponiveis/8/8139/tde-06042021-192617/publico/2020_AndressaVieiraESilva_VCorr.pdf). Acesso em: 14 jul. 2021.

SILVA, Tiago Carvalho da; BARION, Michele Cristiani. **Aplicativo para apresentação de boletim escolar com extração de dados do SUAP através do Web Scraping: estudo de caso no IFSP**. 2019. 24 f. TCC (Graduação) - Curso de Análise e Desenvolvimento de Sistemas, Instituto Federal de Educação, Ciência e Tecnologia de São Paulo, São Paulo, 2019. Disponível em:

[https://hto.ifsp.edu.br/portal/images/thumbnails/images/IFSP/Cursos/Coord\\_ADS/Arquivos/TCCs/2019/TCC\\_Tiago\\_Carvalho\\_da\\_Silva.pdf](https://hto.ifsp.edu.br/portal/images/thumbnails/images/IFSP/Cursos/Coord_ADS/Arquivos/TCCs/2019/TCC_Tiago_Carvalho_da_Silva.pdf). Acesso em: 30 jul. 2021.