

**CENTRO PAULA SOUZA
FACULDADE DE TECNOLOGIA DE FRANCA
“Dr. THOMAZ NOVELINO”**

TECNOLOGIA EM ANÁLISE E DESENVOLVIMENTO DE SISTEMAS

VITOR SANTOS LOPES

TO-POINT

A tecnologia da informação como apoio para apresentar pontos
essenciais de um conteúdo escrito

Trabalho de Graduação
apresentado à Faculdade de Tecnologia de
Franca - “Dr. Thomaz Novelino”, como parte
dos requisitos obrigatórios para obtenção do
título de Tecnólogo em Análise e
Desenvolvimento de Sistemas.

Orientadora: Prof^ª. Ma. Maria Luísa Cervi Uzun

FRANCA/SP

2022

TO-POINT

Vitor Santos Lopes¹

Resumo

Este trabalho tem como objetivo registrar o desenvolvimento de uma aplicação utilizada para gerar um resumo, uma nuvem de palavras de textos e outras fontes de leitura. É apresentado toda a estrutura do projeto, bem como ferramentas utilizadas para a criação da aplicação e o fluxo de trabalho do *software*, que tem como público-alvo pessoas que não tem tempo para fazer a leitura de algum artigo ou livro, mostrando somente as partes essenciais do conteúdo. A metodologia utilizada é a abordagem qualitativa, de natureza aplicada e para atender aos objetivos utilizou-se uma pesquisa descritiva e explicativa através de um procedimento técnico chamado levantamento bibliográfico. A principal ferramenta utilizada nesse projeto é o processamento de linguagem natural com técnica de sumarização de textos para transformar a linguagem natural em informações manipuláveis por máquinas. Pode-se concluir que, o *software* possibilita o público-alvo uma melhor capacidade de extração de texto, filtrando somente as informações importantes.

Palavras-chave: Inteligência Artificial. Nuvem de palavras. PLN. Resumo.

Abstract

The objective of this document is to record the development an application for generate summary and wordcloud from texts and other sources of reading. The entire structure of the project is presented, as well as the tools used to create the application and the workflow of the software, whose target audience is people who do not have time to read an article or book, showing only the parts content essentials. The methodology used is a qualitative approach, applied in nature and to meet the objectives a descriptive and explanatory research was used through a technical procedure called bibliographical survey. The main tool used in this project is natural language processing with a text summarization technique to transform natural language into information that can be manipulated by machines. It can be concluded that the software provides the target audience with a better ability to extract text, filtering only the important information.

Keywords: Artificial intelligence. Word Cloud. NLP. Summary.

1 Introdução

¹ Graduando em Análise e Desenvolvimento de Sistemas pela Fatec Dr Thomaz Novelino – Franca/SP. Endereço eletrônico: vitorsantoslopes1@gmail.com

Atualmente grandes volumes de dados são gerados a cada instante em aplicativos móveis, sistemas, *IoT* (Internet das Coisas), *TVs*, entre outros dispositivos; assim são desenvolvidas técnicas para capturar e processar esses dados, transformando-os em informações que podem ser utilizadas para tomada de decisão.

Grande parte dos dados gerados atualmente são não estruturados, por exemplo: fotos, vídeos, posts, comentários e áudios. Logo foi necessário desenvolver algoritmos para transformar esses dados não estruturados em algum tipo de informação.

No início do ano 2000 a área de inteligência artificial começou a ganhar espaço devido ao avanço da tecnologia, computadores mais sofisticados foram fabricados e uma grande quantidade de dados estavam começando a ser gerados pois a *internet* estava ficando mais acessível às pessoas. De acordo com, Gomes (2010, p. 235):

Quando se fala de Inteligência Artificial, é difícil defini-la, mas ao longo do tempo ela seguiu quatro linhas de pensamento: I. Sistemas que pesam como seres humanos: “O novo e interessante esforço para fazer os computadores pensarem... máquinas com mentes, no sentido total e literal”. (HAUGELAND, 1985). II. Sistemas que atuam como seres humanos: “A arte de criar máquinas que executam funções que exigem inteligência quando executadas por pessoas.” (KURZWEIL, 1990). III. Sistemas que pensam racionalmente: “O estudo das faculdades mentais pelo seu uso de modelos computacionais.” (CHARNIAK; MCDERMOTT, 1985). IV. Sistemas que atuam racionalmente: “A Inteligência Computacional é o estudo do projeto de agentes inteligentes.” (POOLE et al., 1998).

Com o surgimento das *GPUs* (*Graphics Processing Unit*), que realizam processamento de forma paralela, permitiu que surgisse o que é chamado de “Tempestade Perfeita”.

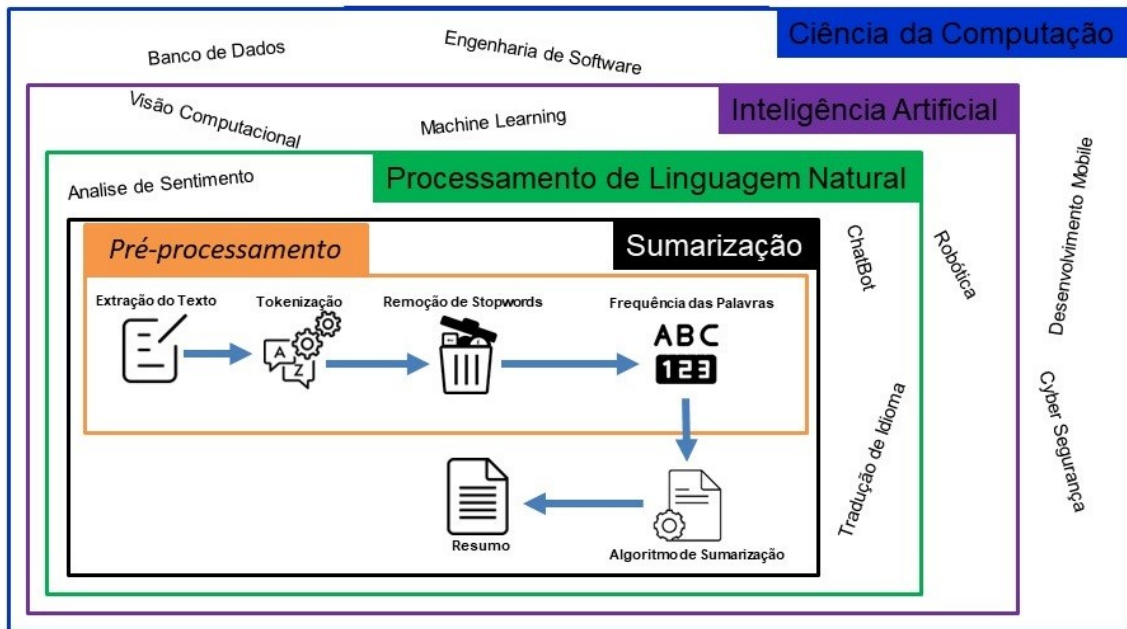
As unidades de processamento gráfico, que permitem realizar operações matemáticas de forma paralela, principalmente operações com matrizes e vetores, elementos presentes em modelos de redes neurais artificiais, formaram a tempestade perfeita, que permitiu a evolução na qual nos encontramos hoje: Big Data + Processamento Paralelo + Modelos de Aprendizagem de Máquina = Inteligência Artificial. (DATA SCIENCE ACADEMY, 2022, p. 1)

Processamento de Linguagem Natural (PLN) é uma subárea da Inteligência Artificial que tem como objetivo tornar os computadores capazes de entender a linguagem humana. Apesar do PLN não ser uma ciência nova, essa tecnologia teve um grande avanço graças aos conjuntos de dados massivos e com estrutura grande chamados de *BigData*.

O objetivo do PLN é fornecer aos computadores a capacidade de entender e compor textos. “Entender” um texto significa reconhecer o contexto, fazer análise sintática, semântica, léxica e morfológica, criar resumos, extrair informação, interpretar os sentidos, analisar sentimentos e até aprender conceitos com os textos processados. (RODRIGUES; JESSICA, 2022, p. 1)

A tradução automática é considerada por muitos como o marco inicial do estudo em línguas naturais utilizando computadores, atualmente a tecnologia de processamento de linguagem natural é bastante utilizada para identificar notícias falsas, por exemplo.

Figura 1 – Etapas de sumarização



Fonte: o autor (2022)

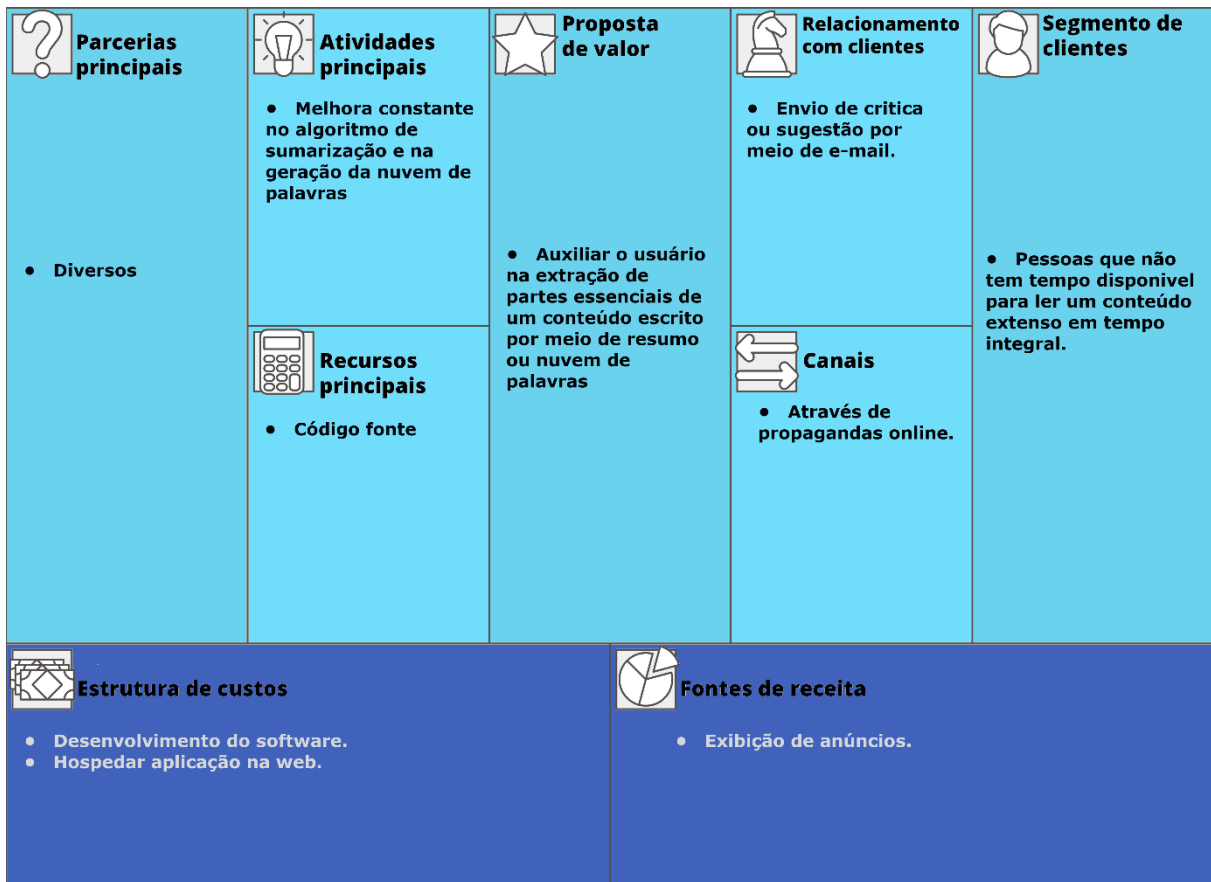
Na figura 1 mostra as etapas de como é realizado a sumarização, o PLN é uma subárea de inteligência artificial que é uma das áreas da ciência da computação. A sumarização é um dos métodos e pode ser utilizado em

processamento de linguagem natural para realizar a extração de pontos importantes tratados no texto. A primeira etapa para realizar uma sumarização é extrair um texto, como artigos, livros e notícias.

Após a extração do texto que se quer resumir, entra a segunda etapa, a *tokenização*, essa etapa simplesmente separa todas as palavras do texto. A próxima etapa é a remoção de *stopwords*, essas são palavras que podem ser consideradas irrelevantes para o conjunto de palavras, exemplos de *stopwords* são: as, e, os, de, para. Logo em seguida vem a quarta etapa, a contagem de com que frequência as palavras aparecem no texto, então é dado um peso para tais palavras dando um significado inicial de sua importância no texto, para que estejam prontos para a quinta etapa, que é a utilização de um algoritmo que calcule quais sentenças são mais relevantes no texto com base nas palavras extraídas, gerando o resumo com as sentenças mais importantes, esta seria a última etapa. A ideia é entregar um produto que possa ajudar estudantes, professores e outros profissionais a entender os pontos principais que estão sendo discutidos em artigos, livros, entre outros documentos em um curto espaço de tempo.

2 Viabilidade do projeto

Figura 2 – Canvas To-Point



Fonte: o autor (2022)

Na figura 2 pode ser visualizada a estratégia de negócio que é utilizada, o objetivo é criar uma ferramenta que atenda às necessidades dos leitores que leem nos pequenos intervalos de tempo livre e que desejam ler somente os principais pontos de um conteúdo. Foi realizado um orçamento inicial e o projeto não tem grandes custos, tendo em vista o tamanho do projeto, é necessário somente o pagamento da hospedagem da aplicação em alguma plataforma e adquirir alguns cursos de extensão para auxiliar no desenvolvimento do *software*. Como fonte de receita, é utilizado a exibição de anúncios na tela quando o conteúdo já foi processado, e para receber o *feedback* dos usuários em relação a aplicação *web*, o usuário pode enviar um *e-mail* para o criador da aplicação por meio da opção de *feedback*.

3 Levantamento de Requisitos

Esta etapa é o início para toda a atividade de desenvolvimento de *software*.

O levantamento de requisitos é uma das partes mais importantes do processo que resultará no desenvolvimento de um sistema. Entender aquilo que o cliente deseja ou o que o cliente acredita que precisa e as regras do negócio ou processos do negócio. Isso é o cerne que move essa importante função que faz parte da engenharia de requisitos. (VERISSIMO; RICARDO, 2022, p. 1)

3.1 Elicitação e especificação dos Requisitos

O levantamento dos requisitos foi realizado uma pesquisa utilizando a técnica de etnografia. De acordo com Somerville (2003, p. 89):

Etnografia é uma técnica de observação que pode ser usada para compreender os processos operacionais e ajudar a extrair os requisitos de apoio para esses processos.

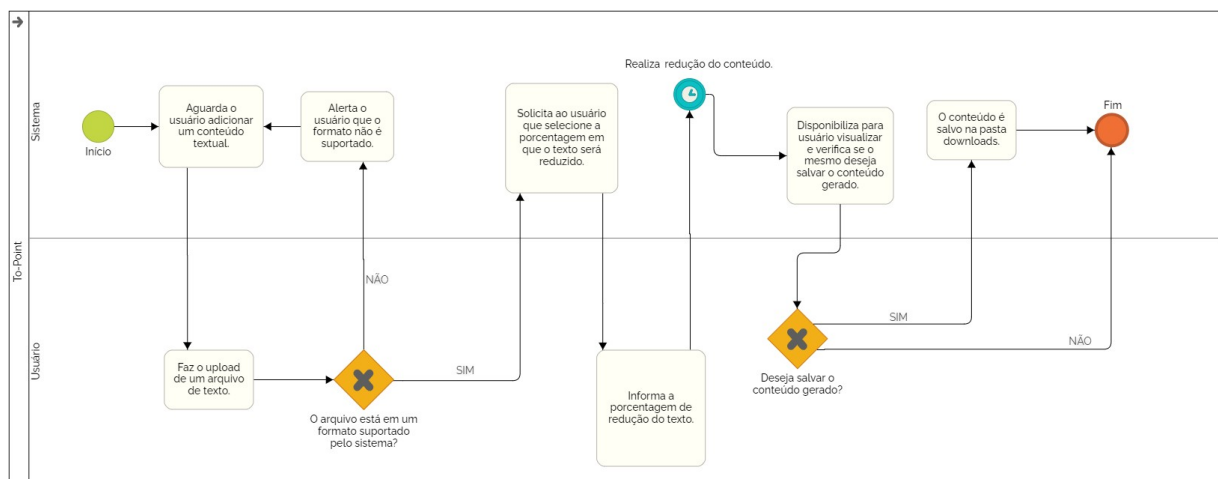
Esta é a primeira fase no ciclo de desenvolvimento de um projeto, nessa fase são definidos o escopo do projeto e as funcionalidades.

3.2 BPMN

O BPMN (*Business Process Model and Notation*) é uma notação que usa a representação gráfica utilizando ícones para simbolizar as etapas de um processo.

Peça-chave na gestão de processos de negócios, representa de forma visual uma sequência detalhada de atividades de negócios e fluxos de informação necessários para concluir um processo. (LUCIDCHART, 2022, p. 1)

Figura 3 – BPMN To-Point



De acordo com o BPMN acima, quando o *software* é iniciado, ele aguarda o usuário realizar um *upload* ou colocar o *link* da página *web* do conteúdo, caso haja algum problema em ler a formatação do arquivo, o usuário é informado. Em seguida o usuário irá informar o nível de redução do conteúdo, então é dado início ao processamento do resumo e da nuvem de palavras. Depois que o novo conteúdo é gerado o usuário poderá visualizar o texto resumido e a nuvem de palavras, ambos estarão disponíveis para serem salvos na máquina do usuário.

3.3 Requisitos Funcionais

Os requisitos funcionais indicam todos os processos que o sistema deve fazer.

Fonte: o autor (2022)

Quadro 1 – Requisitos Funcionais do sistema

RF001 -Armazenar Conteúdo Original	Categoria: <input checked="" type="checkbox"/> Oculto <input type="checkbox"/> Evidente	Prioridade: <input checked="" type="checkbox"/> Altíssima <input type="checkbox"/> Alta <input type="checkbox"/> Média <input type="checkbox"/> Baixa
Descrição: O sistema deve salvar o conteúdo original em uma variável na memória até que o <i>software</i> seja encerrado ou voltar na tela de início.		
RF002 -Armazenar do Resumo	Categoria: <input checked="" type="checkbox"/> Oculto <input type="checkbox"/> Evidente	Prioridade: <input checked="" type="checkbox"/> Altíssima <input type="checkbox"/> Alta <input type="checkbox"/> Média <input type="checkbox"/> Baixa
Descrição: O sistema deve salvar o resumo em uma variável na memória até que o <i>software</i> seja encerrado ou voltar na tela de início.		
RF003 -Armazenar nuvem de palavras	Categoria: <input checked="" type="checkbox"/> Oculto <input type="checkbox"/> Evidente	Prioridade: <input checked="" type="checkbox"/> Altíssima <input type="checkbox"/> Alta <input type="checkbox"/> Média <input type="checkbox"/> Baixa
Descrição: O sistema deve salvar a nuvem de palavras em uma variável na memória até que o <i>software</i> seja encerrado ou voltar na tela de início.		
RF004 -Descartar variável redução	Categoria: <input checked="" type="checkbox"/> Oculto <input type="checkbox"/> Evidente	Prioridade: <input type="checkbox"/> Altíssima <input type="checkbox"/> Alta <input checked="" type="checkbox"/> Média <input type="checkbox"/> Baixa
Descrição: O sistema deve descartar a variável de porcentagem de redução quando o resumo do conteúdo for processado.		
RF005 -Processar Nuvem de Palavras	Categoria: <input checked="" type="checkbox"/> Oculto <input type="checkbox"/> Evidente	Prioridade: <input type="checkbox"/> Altíssima <input type="checkbox"/> Alta

		(X) Média () Baixa
Descrição: O sistema deve construir a nuvem de palavras após o resumo ser processado.		
RF006-Validar Arquivo ou Link	Categoria: () Oculto (X) Evidente	Prioridade: () Altíssima (X) Alta () Média () Baixa
Descrição: O sistema deve validar o arquivo ou <i>link</i> passado pelo usuário antes de processá-lo.		
RF007-Exibir Anúncios	Categoria: () Oculto (X) Evidente	Prioridade: () Altíssima () Alta (X) Média () Baixa
Descrição: O sistema deve exibir um anúncio após o processamento do conteúdo.		
RF008-Construir Nuvem de Palavras	Categoria: () Oculto (X) Evidente	Prioridade: () Altíssima () Alta () Média (X) Baixa
Descrição: O sistema deve utilizar as palavras mais frequentes no conteúdo, para construir a nuvem de palavras.		

3.4 Requisitos Não Funcionais

Os requisitos não funcionais descrevem todas as características do sistema.

Quadro 2 – Requisitos Não Funcionais do Sistema

RNF001- Utilizar <i>links</i> externos	O usuário pode utilizar um <i>link</i> de um texto para o sistema processar.	Tipo: Externo	() Desejável (X) Obrigatório	(X) Permanente () Transitório
RNF002- Lingua	Só serão processados conteúdo em português-Brasil.	Tipo: Desempenho	(X) Desejável () Obrigatório	() Permanente (X) Transitório
RNF003- Disponibilidade	O sistema deve estar disponível somente em aplicação <i>web</i> .	Tipo: Produto	() Desejável (X) Obrigatório	() Permanente (X) Transitório
RNF004- Formato do arquivo de upload	O usuário pode fazer o upload de um arquivo no formato <i>TXT</i> ou <i>PDF</i> , somente.	Tipo: Usabilidade	(X) Desejável () Obrigatório	() Permanente (X) Transitório
RNF005- Níveis de redução	O sistema deverá ter três níveis de redução de texto: baixo remove 25% do conteúdo original, médio remove 50% e alto remove 75%.	Tipo: Usabilidade	() Desejável (X) Obrigatório	(X) Permanente () Transitório
RNF006- Canal de <i>feedback</i>	O usuário pode fazer críticas ou sugestões por meio do botão <i>feedback</i> localizado na parte de baixo da página.	Tipo: Interface	(X) Desejável () Obrigatório	() Permanente (X) Transitório
RNF007-	O resumo pode ser	Tipo: Externo	() Desejável	() Permanente

Exportar resumo	exportado somente no formato <i>PDF</i> .		(X) Obrigatório	(X) Transitório
RNF008- Exportar nuvem de palavras	A nuvem de palavras poderá ser exportada somente no formato <i>PNG</i> .	Tipo: Externo	() Desejável (X) Obrigatório	() Permanente (X) Transitório
RNF009- Cor principal	A cor predominante no <i>software</i> deve ser a cinza-claro.	Tipo: Interface	(X) Desejável () Obrigatório	(X) Permanente () Transitório

3.5 Regras de Negócio

As regras de negócio são essenciais para o funcionamento do sistema pois ajudam a regular todas as operações dele.

Quadro 3 – Regras de Negócio do Sistema.

RN001 – Acesso do usuário
Descrição: Qualquer usuário tem acesso total aos recursos do <i>software</i> .
RN002 – Similaridade do Conteúdo
Descrição: Independente da fonte do conteúdo, se o mesmo for similar a outro, deverá ser gerado uma saída similar para ambos.

3.6 Casos de Uso

Os casos de uso proporcionam uma visão de como o sistema irá se comportar.

Índice de casos de uso:

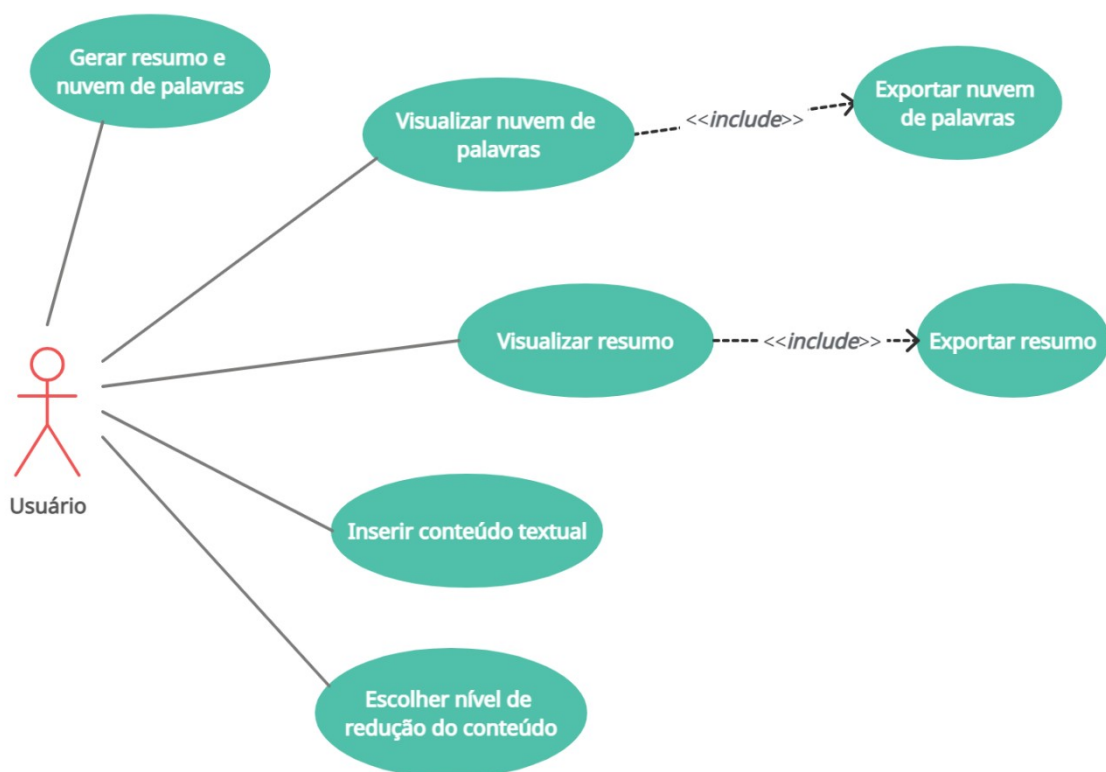
- UC 001: Inserir conteúdo textual
- UC 002: Escolher nível de redução do conteúdo
- UC 003: Visualizar resumo
- UC 004: Visualizar nuvem de palavras
- UC 005: Gerar resumo e nuvem de palavras

Diagrama de casos de uso:

A Figura 4 representa o diagrama de casos de uso do sistema.

O diagrama de caso de uso resume os detalhes dos usuários do seu sistema (também conhecidos como atores) e as interações deles com o sistema. (LUCIDCHART, 2022, p. 1)

Figura 4 – Diagrama de casos de uso do sistema



Fonte: o autor (2022)

Especificação dos casos de uso:

Quadro 4 – Caso de uso inserir conteúdo textual

Caso de Uso – Inserir conteúdo textual	
ID	UC 001
Descrição	Este caso de uso tem por objetivo permitir que o usuário insira um arquivo no formato <i>TXT</i> , <i>PDF</i> ou um <i>link</i> de uma página.
Ator Primário	Usuário
Pré-condição	Nenhuma
Cenário Principal	<ol style="list-style-type: none"> 1. O use case inicia quando o usuário faz o upload de um arquivo ou colocar o <i>link</i> de alguma página. 2. O sistema verificar o formato do arquivo e/ou a página web possui algum texto. 3. O sistema confirma se a importação dos dados foi realizada com sucesso.
Pós-condição	Não possui
Cenário Alternativo	<p>2a – O arquivo não é do formato <i>TXT</i>, <i>PDF</i> ou uma página <i>WEB</i>.</p> <p>2a.1 O sistema informa que ocorreu um erro na importação.</p> <p>2a.2 Retorna ao passo 1 do cenário principal.</p> <p>2b – Não foi encontrado texto na página <i>WEB</i>.</p> <p>2b.1 O sistema avisa que a página informada não é válida.</p> <p>2b.2 Retorna ao passo 1 do cenário principal.</p>

Quadro 5 – Caso de uso escolher nível de redução do conteúdo

Caso de Uso – Escolher nível de redução do conteúdo

ID	UC 002
Descrição	Este caso de uso tem por objetivo solicitar que o usuário selecione o nível de redução que é aplicado no conteúdo.
Ator Primário	Usuário
Pré-condição	Nenhuma
Cenário Principal	<ol style="list-style-type: none"> 1. O usuário seleciona a opção desejada para reduzir o conteúdo 2. O sistema deve verificar se o usuário adicionou o conteúdo ao sistema.
Pós-condição	Não possui
Cenário Alternativo	<p>2a – O usuário não adicionou o conteúdo ao sistema.</p> <p>2a.1 O sistema informa que o usuário deve adicionar o conteúdo que vai ser resumido.</p>

Quadro 6 – Caso de uso visualizar resumo

Caso de Uso – Visualizar resumo	
ID	UC 003
Descrição	Este caso de uso tem por objetivo visualizar o resumo gerado pelo sistema.
Ator Primário	Usuário
Pré-condição	Nenhuma
Cenário Principal	<ol style="list-style-type: none"> 1. O use case inicia quando o sistema finaliza o processamento do conteúdo. 2. O usuário é informado que o processamento foi finalizado. 3. O conteúdo resumido é exibido.
Pós-condição	Não possui
Cenário Alternativo	Não possui

Quadro 7 – Caso de uso visualizar nuvem de palavras

Caso de Uso – Visualizar nuvem de palavras	
ID	UC 004
Descrição	Este caso de uso tem por objetivo visualizar a nuvem de palavras gerada pelo sistema.
Ator Primário	Usuário
Pré-condição	Nenhuma
Cenário Principal	<ol style="list-style-type: none"> 1. O use case inicia quando o sistema finaliza o processamento do conteúdo. 2. O usuário é informado que o processamento foi finalizado. 3. O usuário seleciona a opção nuvem de palavras. 4. A nuvem de palavras é exibida.
Pós-condição	Não possui
Cenário Alternativo	Não possui

Quadro 8 – Caso de uso gerar resumo e nuvem de palavras

Caso de Uso – Gerar resumo e nuvem de palavras	
ID	UC 005
Descrição	Este caso de uso tem por objetivo gerar o resumo e a nuvem de palavras
Ator Primário	Usuário
Pré-condição	Nenhuma
Cenário	<ol style="list-style-type: none"> 1. O use case inicia quando o usuário seleciona a opção gerar.

Principal	<ol style="list-style-type: none"> 2. O sistema verificar se os campos de conteúdo e nível de redução foram preenchidos. 3. O sistema gera o resumo do conteúdo original e a nuvem de palavras.
Pós-condição	Não possui
Cenário Alternativo	<ol style="list-style-type: none"> 2a – O usuário não preencheu algum campo 2a.1 O sistema alerta o usuário que todos os campos devem estar preenchidos e não prossegue até os campos estarem completos.

4 Ferramentas e Métodos ou Desenvolvimento

Após realizar uma pesquisa sobre quais ferramentas utilizar no desenvolvimento do projeto, chegou-se à conclusão de que a utilização das ferramentas descritas abaixo foi uma escolha que envolve várias tecnologias.

4.1 Ferramentas

Grande parte da aplicação foi desenvolvida utilizando a linguagem de programação criada por Guido Van Rossum, o *Python*, de acordo com o site *DevMedia*: “*Python* traz características que possibilitam escrever o mesmo requisito em menos linhas de código que o necessário em outras linguagens de programação” (DEVMEDIA, 2022, p.1).

No *front-end* foi utilizado *html*, *css* e *javascript* juntamente com o *framework Bootstrap*.

Para versionamento do projeto foram utilizados *Git* e *Github*. O *Git* é um sistema utilizado para controle de versão, ele cria um histórico de alterações que foram realizadas no projeto, podendo até reverter para uma versão anterior no histórico. Já o *Github* é um serviço que hospeda repositórios do *Git*, permitindo a colaboração em um projeto pois o *Github* hospeda os repositórios remotamente (GITHUB, 2022).

Foram utilizadas bibliotecas *Spacy*, *Sumy* e *NLTK* com *Python*, para realizar o processamento e sumarização do texto e a geração da nuvem de palavras, *NLTK* é uma das bibliotecas mais poderosas no mercado quando se fala em processamento de linguagem natural, segundo o site oficial “*NLTK* é uma plataforma líder para construir programas *Python* para trabalhar com dados de linguagem humana.” (NLTK, 2022, p.1). *Spacy* são bibliotecas de processamento avançado de linguagem natural, construídas para serem utilizadas em produtos reais, suporta treinamento de

modelos de processamento de linguagem natural para mais de 70 idiomas. Já o *Sumy* foi criado inicialmente a partir de uma tese que tinha como necessidade a redução do tamanho dos artigos tcheco/eslovaco. As três bibliotecas citadas são de código aberto e sem custos para uso.

E por fim, a aplicação foi disponibilizada em uma plataforma web que tem como especialidade hospedar aplicações em *Python* chamada *Pythonanywhere*. A plataforma utiliza container *Docker* para armazenar a aplicação, pois com *Docker* é garantido o isolamento dos recursos utilizados, garantindo que funcione independentemente do sistema que a aplicação está rodando. O registro do domínio, que é um endereço único para um site, foi contratado um plano do *Registro.br* para utilizar o domínio `<http://www.topoint.dev.br/>`.

4.2 Métodos e Ferramentas

Para o desenvolvimento da aplicação, a principal ferramenta utilizada é o processamento de linguagem natural com técnica de sumarização de textos, para transformar a linguagem natural em informações manipuláveis por máquinas.

Para isso, primeiramente foi realizado o levantamento dos requisitos e a construção da documentação com todas as características do *software*, garantindo que ele atenda o seu público-alvo. Em seguida foi criado um cronograma com as etapas da construção do sistema, estruturando o fluxo de trabalho de desenvolvimento.

Na sequência foi feita uma pesquisa para descobrir os recursos que seriam utilizadas no projeto e o conhecimento para utilizá-los.

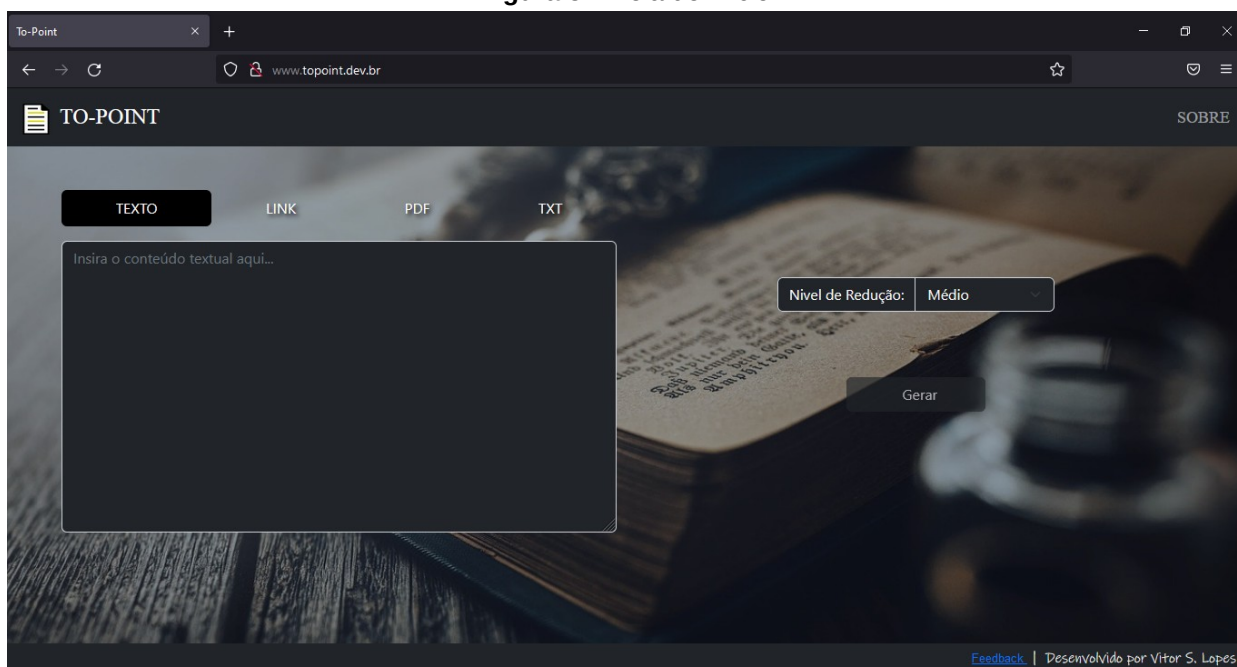
Após a escolha das ferramentas, deu-se início a implementação do algoritmo de sumarização. O algoritmo utilizado foi o *LexRank* combinado com o *TextRank*, são algoritmos inspirados no *world wide web*. Eles tentam encontrar conexões entre as frases e identificar aquelas relacionadas com as palavras mais significativas.

O próximo passo, foi criar o *layout* da aplicação, deixando o sistema de acordo com o que foi escrito na documentação. Por fim, a interface foi integrada com os algoritmos utilizados para o funcionamento da aplicação *web*.

5 Resultados e Discussão

Na figura 5, pode ser visto a tela de início da aplicação e as opções disponíveis para o usuário. Inicialmente insere-se um *link* de uma página *web* ou um arquivo *PDF* ou *TXT* que contenha um conteúdo textual, na sequência, vai aparecer os níveis de redução para serem aplicados no texto e a opção de dar feedback na parte de baixo da tela. O botão “sobre” é para o usuário conhecer um pouco sobre do que se trata o projeto, e por fim, a tecla gerar que irá iniciar o processo de redução do texto.

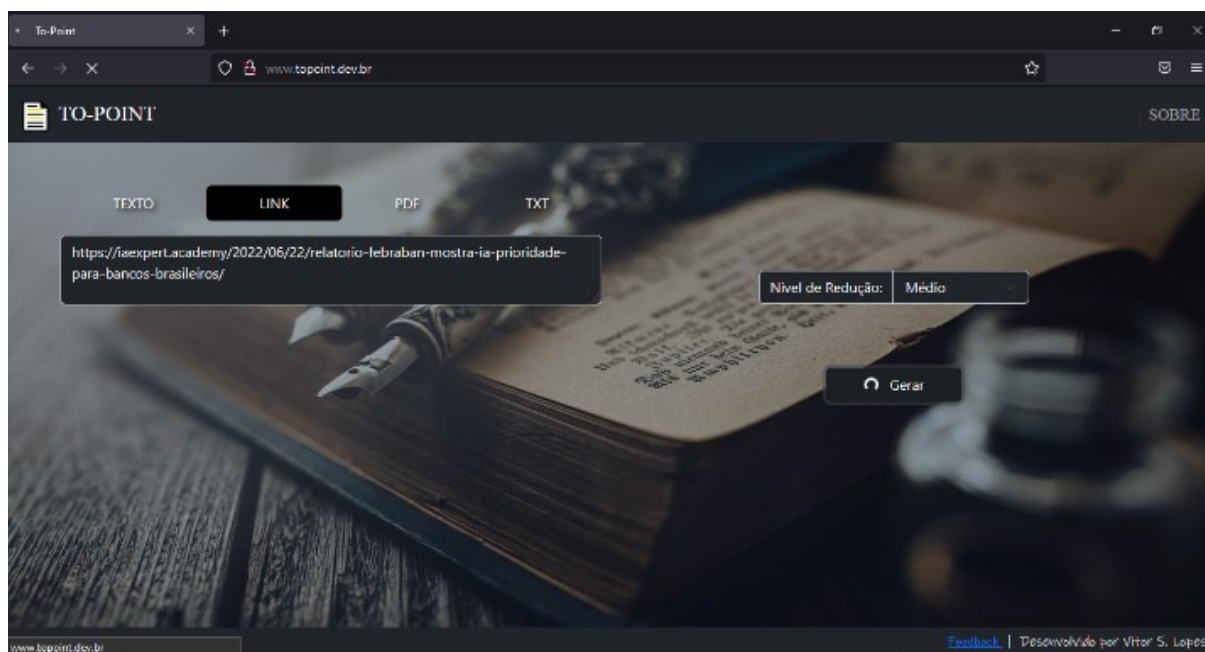
Figura 5 – Tela de início



Fonte: o autor (2022)

Na figura 6, o usuário inseri um *link* de uma página *web*, depois seleciona o nível de redução que o conteúdo é reduzido, neste caso, o nível selecionado foi o médio, que inclusive é o nível padrão, em seguida o usuário clicou no botão “Gerar” e então aparece um ícone de carregamento, indicando que o conteúdo está sendo validado pelo sistema, iniciando o processo de geração do resumo.

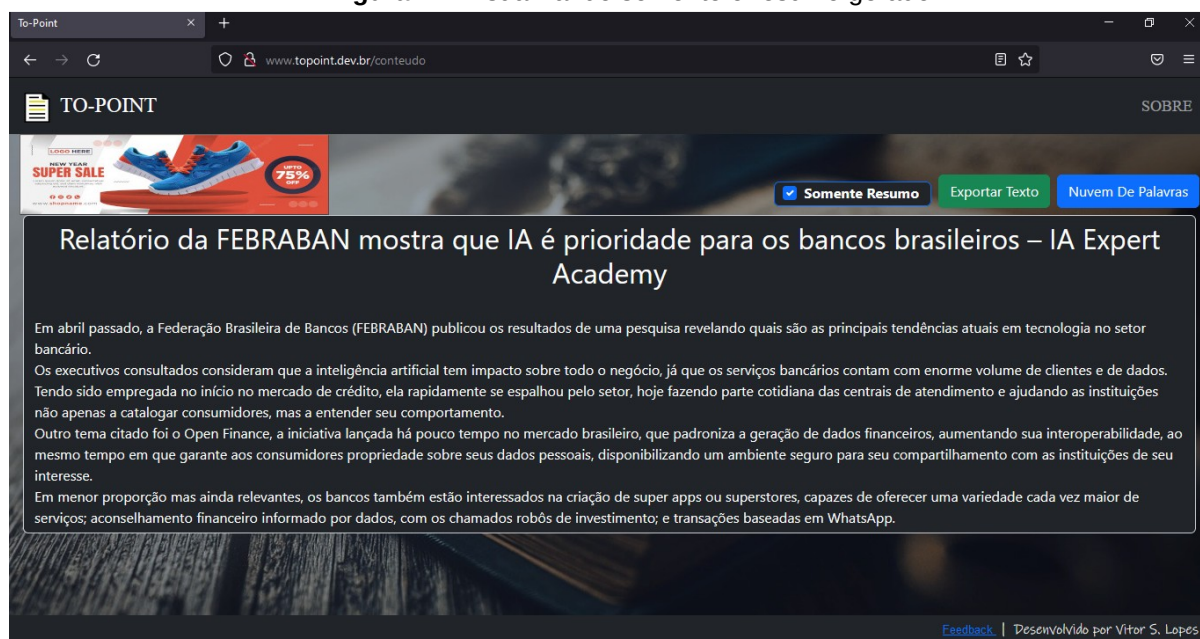
Figura 6 – O sistema está processando o conteúdo



Fonte: o autor (2022)

A figura 7 aponta onde o resumo é exibido, por esse motivo a opção “Somente Resumo” está marcada. Ela irá mostrar somente a parte do texto que o algoritmo de processamento natural julgou importante. É possível ver também que existe um anúncio na parte superior da tela, abaixo do logo da aplicação, anúncios são gerados aleatoriamente para que a aplicação retorne algum tipo de receita.

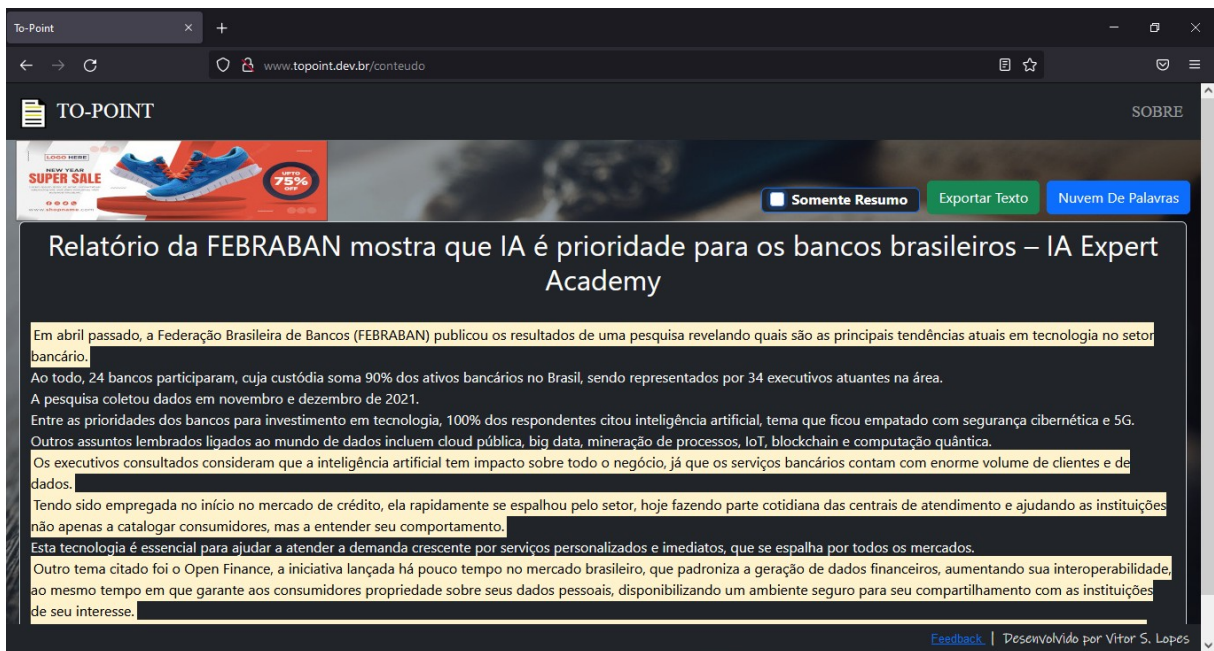
Figura 7 – Visualizando somente o resumo gerado



Fonte: o autor (2022)

Na figura 8 tem-se a mesma tela de visualização do resumo mostrado na figura anterior, a diferença é que a opção “Somente Resumo” está desmarcada, significa que o sistema irá mostrar o conteúdo original extraído do arquivo ou site e, em realce, irá mostrar as principais sentenças escolhidas pelo algoritmo de processamento de linguagem natural. Existe acima do resumo gerado, um botão com o nome “Exportar Texto”, a função desse botão é exportar o conteúdo extraído para o formato *PDF*.

Figura 8 – Tela de visualizar o resumo com o conteúdo original

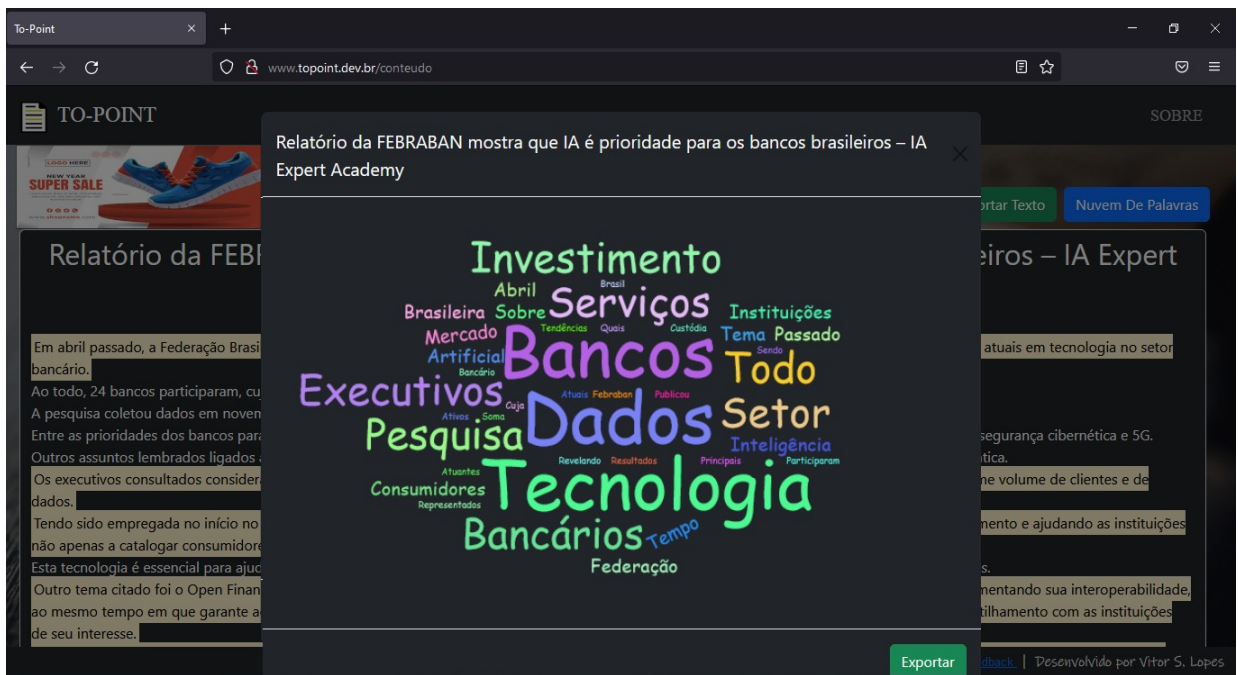


A figura 9 é a tela para visualizar a nuvem de palavras. Essa janela é aberta quando o usuário clica no botão “Nuvem de Palavras”, desta forma, é mostrado para

Fonte: o autor (2022)

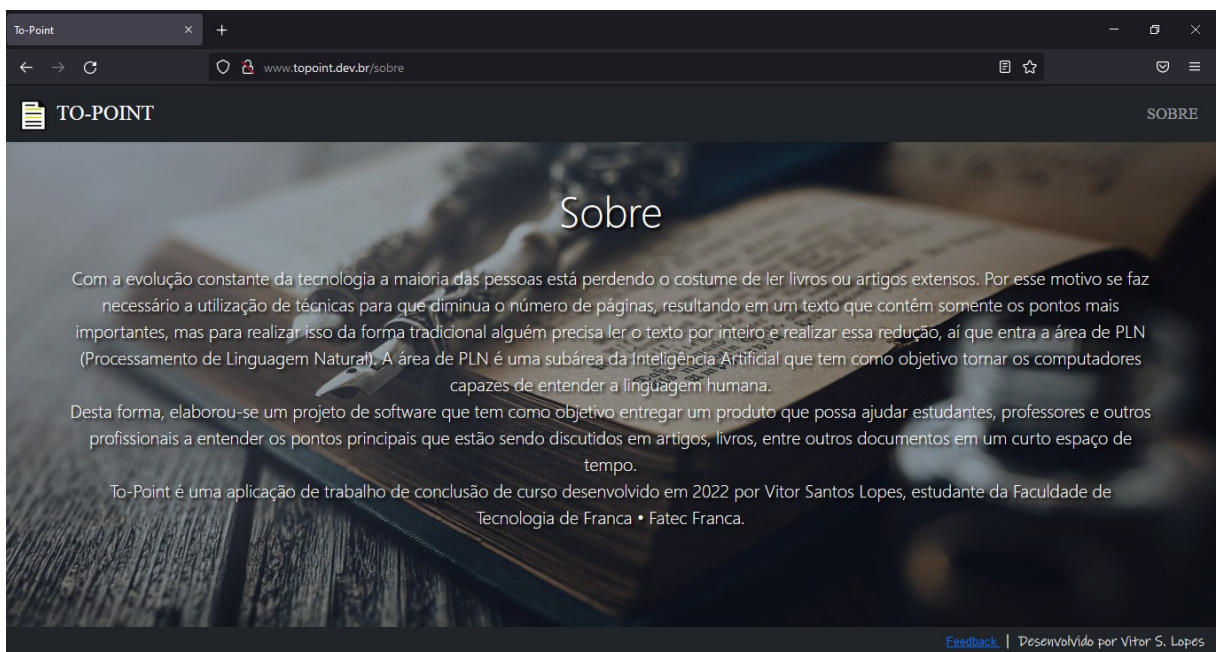
o usuário as palavras-chave do conteúdo escrito, o tamanho da palavra indica a sua importância no texto. Na parte de baixo da janela existe um botão chamado “Exportar”, ele exporta a nuvem de palavras no formato *PNG*.

Figura 10 – Tela onde é falado sobre o projeto



Fonte: o autor (2022)

A figura 10 tem-se uma introdução falando a função do *software* e objetivos do projeto, logo abaixo, ao lado do nome do desenvolvedor que projetou o *software*, tem a opção de “*Feedback*”, é onde o usuário do sistema pode deixar críticas e sugestões. Todo *feedback* é enviado para o e-mail do desenvolvedor do *software*.



Fonte: o autor (2022)

Considerações finais

O objetivo inicial do projeto foi aplicar algoritmos de aprendizado de máquina em algum *software* tornando um produto utilizando técnicas de processamento de linguagem natural e sumarização para a extração do texto e para gerar numa nuvem de palavras facilitando a compreensão de determinado assunto que o usuário passar como entrada para a aplicação.

A primeira ideia *To-Point* iria ser um aplicativo somente para dispositivos móveis Android, mas o desenvolvimento se tornou inviável pois o intuito era colocar todo o processamento da aplicação no dispositivo móvel e alguns algoritmos não teriam uma boa performance, então o software foi refeito quase que completamente para a plataforma web, pois tem vantagens como o acesso de qualquer dispositivo e não utiliza muitos recursos do usuário que está navegando na página web.

Perto do fim de desenvolvimento da aplicação, novos conhecimentos foram adquiridos e alguns requisitos tiveram que ser alterados para a entrega de um produto minimamente viável, também foi criado um cronograma para o gerenciamento do projeto, aumentando as chances do projeto ser finalizado.

O resultado final do projeto foi muito satisfatório, pois o propósito inicial da aplicação foi cumprido com sucesso. Por mais que o projeto tenha alcançado um bom resultado, ainda existe possibilidade de melhoria com novas funcionalidades como, seleção de outras opções de algoritmos de sumarização, tamanho e cor da fonte, cor da nuvem de palavras, e otimizações de processamento.

O link do software desenvolvido é: <<http://www.topoint.dev.br/>>.

O link da página web utilizada na apresentação é: <<https://iaexpert.academy/2022/09/14/ia-identifica-vocalizacoes-roedores-usam-para-comunicar/>>.

O link do arquivo PDF utilizado na apresentação é: <<https://periodicos.ifsul.edu.br/index.php/thema/article/view/1104/1152>>.

Referências

DATA SCIENCE ACADEMY, **Capítulo 1 – Deep Learning e a Tempestade Perfeita**. Disponível em: < <https://www.deeplearningbook.com.br/deep-learning-a->

tempestade-perfeita/>. Acesso em: 26. agosto, 2022.

DEV MEDIA, **Guia Completo de Python**. Disponível em: <<https://www.devmedia.com.br/guia/python/37024>>. Acesso em: 26. março, 2022.

DOCKER, **DOCKER**. Disponível em: <<https://www.docker.com/>>. Acesso em: 27. novembro, 2022.

FIA, **Big Data: como funciona, exemplos, importância e desafios**. Disponível em: <<https://fia.com.br/blog/big-data/>>. Acesso em: 26. agosto, 2022.

GITHUB. Disponível em: <<https://github.com/>>. Acesso em: 27. agosto, 2002.

GOMES, D. dos S. Inteligência Artificial: conceitos e aplicações. **Olhar Científico**. v1, n. 2, p. 234-246, 2010.

LUCIDCHART, **Diagrama de caso de uso UML**. Disponível em: <<https://www.lucidchart.com/pages/pt/diagrama-de-caso-de-uso-uml>>. Acesso em: 05. março, 2022.

LUCIDCHART, **O que é BPMN?**. Disponível em: <<https://www.lucidchart.com/pages/pt/o-que-e-bpmn>>. Acesso em: 26. fevereiro, 2022.

MISO-BELICA, **Summarization methods**. Disponível em: <<https://miso-belica.github.io/sumy/summarizators.html>>. Acesso em: 23. maio, 2022.

NLTK, **NLTK**. Disponível em: <<https://www.nltk.org/>>. Acesso em: 27. março, 2022.

PYTHONANYWHERE, **PYTHONANYWHERE**. Disponível em: <<https://www.pythonanywhere.com/>>. Acesso em: 27. novembro, 2022.

REGISTRO.BR, **REGISTRO.BR**. Disponível em: <<https://registro.br/>>. Acesso em: 27. novembro, 2022.

RODRIGUES, Jéssica. **O que é o Processamento de Linguagem Natural?**. Disponível em: <<https://medium.com/botsbrasil/o-que-%C3%A9-o-processamento-de-linguagem-natural-49ece9371cff>>. Acesso em: 26. agosto, 2022.

SAGIROGLU, Seref; SINANC, Duygu. Big data: uma revisão. In: **2013 International Conference on Collaboration Technologies and Systems (CTS)**. IEEE, 2013. p. 42-47.

SAS, **Processamento de Linguagem Natural**. Disponível em: <https://www.sas.com/pt_br/insights/analytics/processamento-de-linguagem-natural.html>. Acesso em: 26. agosto, 2022.

SOMERVILLE, Ian. **Engenharia de software**. 6° ed. Tradução **Maurício de Andrade**. São Paulo: Ed Addison-Wesley, 2003.

SPACY, **SPACY**. Disponível em: < <https://spacy.io/>>. Acesso em: 27. novembro, 2022.

VERISSIMO, Ricardo. **Levantamento de Requisitos e Mapeamento de Processos**. Disponível em: <<http://www.linhadecodigo.com.br/artigo/1564/levantamento-de-requisitos-e-mapeamento-de-processos.aspx>>. Acesso em: 01. março, 2022.