

Chelf: Aplicativo de Busca de Receitas Culinárias Utilizando Técnicas de Mineração de Dados e Processamento de Linguagem Natural

Lucas Carvalho Maracaibe¹, Fabricio Oliveira de Carvalho¹, Abiezer Sidonio Jose Odanaca Junior¹, Matheus Henrique Amorim Moises de Lima¹, Walquiria Fernandes Marins²

¹ Bacharelado em Engenharia de Software – Universidade Evangélica de Goiás – Anápolis – GO – Brazil

² Universidade Evangélica de Goiás – Anápolis – GO – Brazil

{maracaipelucas, ffabricioinformatica, abiezerodanacajosejunior}@gmail.com
, matheushenriquetrab@hotmail.com,
walquiria.marins@docente.unievangelica.edu.br

Abstract. *Cooking at home offers several advantages, such as financial savings and the consumption of healthy meals, and can boost the search for recipes on the internet. This article addresses the development of an application that uses data mining techniques to collect culinary recipes and search for dishes based on the ingredients entered by the user. To achieve satisfactory results, various management methods and development approaches were used. The use of Natural Language Processing (NLP) in the recognition of ingredients from collected recipes, combined with data mining technology, resulted in an efficient and innovative application in the culinary area, as it helps in the discovery of new dishes, in optimizing time in kitchen and the adoption of healthier eating habits.*

Resumo. *Cozinhar em casa oferece diversas vantagens, como a economia financeira e o consumo de refeições saudáveis, e pode impulsionar a busca por receitas na internet. Este artigo aborda o desenvolvimento de uma aplicação que utiliza técnicas de mineração de dados para coletar receitas culinárias e buscar pratos com base nos ingredientes informados pelo usuário. Para atingir resultados satisfatórios, foram utilizados diversos métodos de gerenciamento e abordagens de desenvolvimento. O emprego de Processamento de Linguagem Natural (PLN) no reconhecimento de ingredientes das receitas coletadas, aliado à tecnologia de mineração de dados, resultou em uma aplicação eficiente e inovadora na área culinária, pois auxilia na descoberta de novos pratos, na otimização do tempo na cozinha e na adoção de hábitos alimentares mais saudáveis.*

1. Introdução

O crescimento demográfico, a industrialização e a urbanização geram mudanças significativas no estilo de vida das pessoas, o que resulta em um aumento do sedentarismo e do consumo de alimentos prontos e altamente calóricos (CAMPOS, et al 2023). Essa situação contribui para o aumento de problemas de saúde, como obesidade, hipertensão e certos tipos de câncer (CAMPOLINO, 2022). Diante desse cenário, é crucial promover ações que estimulem uma alimentação saudável, com ênfase em alimentos *in natura* ou minimamente processados (BRASIL MS, 2014). Essas medidas são fundamentais para combater os problemas de saúde relacionados ao estilo de vida atual e para promover o bem-estar da população.

Entretanto, diversas barreiras podem dificultar a preparação de refeições caseiras, como a ausência de modelo familiar, falta de conhecimento, habilidades culinárias limitadas, restrições de tempo e falta de interesse (MAZZONETTO et al., 2020). Com jornadas de trabalho exaustivas e a busca por praticidade, tanto para aqueles que trabalham fora quanto para os que trabalham em casa, a demanda por opções rápidas é alta. Nesse contexto, as plataformas de culinária on-line desempenham um importante papel como fonte de inspiração, especialmente para a geração nascida a partir de 1981, na qual 80% procuram receitas em sites especializados (PORCH, 2022).

As plataformas de culinária on-line, como Tudo Gostoso e Receita Da Hora, desempenham um papel fundamental ao oferecer uma variedade de receitas que atendem às necessidades individuais e tornam a preparação de refeições mais conveniente no contexto agitado do cotidiano. Embora o avanço tecnológico possa contribuir para a ocupação das pessoas e negligência de habilidades culinárias básicas, ele também simplificou o acesso a receitas de qualidade. Essas plataformas se tornaram essenciais para aqueles que buscam opções práticas e saborosas, proporcionando uma solução para a falta de tempo e conhecimento culinário.

Este estudo visa criar uma plataforma on-line, acessível por dispositivos móveis, que centralize receitas e permita aos usuários encontrá-las com base nos ingredientes disponíveis. Essa abordagem soluciona problemas como a necessidade de conhecer previamente o nome da receita, além de facilitar o acesso a opções saudáveis e diversificadas de preparações culinárias. Por meio de algoritmos de Processamento de Linguagem Natural, os usuários poderão buscar receitas que possam ser feitas com os ingredientes que já possuem em casa, tornando a atividade de cozinhar mais prática, eficiente e econômica.

Neste artigo, são aplicadas técnicas avançadas de Mineração de Dados e Processamento de Linguagem Natural (PLN) com o intuito de extrair, identificar e classificar os ingredientes presentes nas receitas disponíveis na internet. Para atingir esse objetivo, são utilizados uma variedade de métodos de gerenciamento e abordagens de desenvolvimento, como *Kanban*, *Scrum* e *Lean Inception*. Além disso, são empregadas práticas de prototipação, utilização de ferramentas como o PM Canva, criação de *personas* e mapeamento da jornada do usuário. Essas estratégias combinadas

forneem uma base s3lida para o sucesso do projeto, permitindo uma an3lise precisa e eficiente dos dados de receitas e a cria33o de recursos avan3ados para os usu3rios.

O objetivo deste estudo foi desenvolver uma aplica33o que utiliza t3cnicas de minera33o de dados para fornecer sugest3es e recomenda3es de receitas com base nos ingredientes informados pelo usu3rio. A aplica33o foi projetada com uma interface do sistema que permite ao usu3rio inserir seus ingredientes e receber recomenda3es personalizadas. Para atingir esse objetivo, foram realizadas revis3o bibliogr3fica e abordagem qualitativa como parte do processo de pesquisa.

Este trabalho est3 organizado da seguinte forma: A Se33o 01, Introdu33o, aborda a motiva33o e o objetivo da pesquisa. A Se33o 02, Referencial Te3rico, oferece uma breve explica33o sobre o assunto que serviu de base para a cria33o do prot3tipo da aplica33o. A Se33o 03, Metodologia, apresenta a problem3tica e os questionamentos que ser3o abordados no estudo. A Se33o 04, intitulada "Chelf", explora as abordagens das tecnologias de PLN e Crawler e sua aplica33o no contexto do trabalho. A Se33o 05, Resultados, apresenta a pesquisa realizada, incluindo os testes com os usu3rios e a discuss3o dos resultados obtidos. Por fim, a Se33o 06, Conclus3o, oferece uma an3lise dos resultados e uma s3ntese das principais descobertas do estudo.

2. Referencial Te3rico

O referencial te3rico deste estudo abrange aspectos relacionados 3 alimenta33o saud3vel, 3s metodologias 3geis de desenvolvimento, 3s tecnologias de minera33o de dados e PLN, e ao *design* e interface de usu3rio. Essa fundamenta33o fornece a base necess3ria para a constru33o do aplicativo proposto e a an3lise dos resultados obtidos. Al3m disso, nesta se33o, s3o exploradas as principais caracter3sticas e limita3es dos estudos relacionados, abordando os conceitos e metodologias relevantes na 3rea, bem como as perguntas a serem respondidas e as ferramentas utilizadas no desenvolvimento do aplicativo.

2.1 Metodologias 3geis

No campo do desenvolvimento de sistemas, as metodologias 3geis s3o abordagens de desenvolvimento de software que priorizam a colabora33o, a flexibilidade e a entrega r3pida de valor. Diferentemente de outras metodologias, como o modelo em cascata, elas s3o iterativas e adaptativas, e permitem que os projetos evoluam de acordo com as necessidades dos usu3rios e as mudan3as no ambiente de neg3cios. Algumas das metodologias 3geis mais populares incluem *Scrum*, *Kanban* e *Lean Inception*.

A Engenharia de *Software* 3 uma 3rea de relev3ncia para o desenvolvimento de sistemas computacionais, pois, nessa 3rea s3o definidas t3cnicas e metodologias que contribuem para a garantia da qualidade do produto final, bem como para a organiza33o da produ33o do sistema. Neste trabalho, destaca-se o uso das metodologias 3geis, em especial as metodologias *Scrum* e XP, com o intuito de aumentar a velocidade de produ33o do *software* e promover a aceita33o 3s mudan3as nos requisitos (ARAG3O, 2022).

No contexto deste projeto, as metodologias ágeis serão utilizadas para orientar o desenvolvimento do aplicativo e garantir que ele seja responsivo às necessidades dos usuários. A aplicação de técnicas ágeis, como prototipação, *sprints* e revisões constantes, permite que a equipe de desenvolvimento identifique rapidamente problemas e oportunidades de melhoria, ajustando o aplicativo conforme necessário para atender às expectativas dos usuários e entregar um produto de qualidade, além de garantir a documentação de decisões e organizações da comunicação do projeto.

2.2 Processamento de Linguagem Natural (PLN)

O Processamento de Linguagem Natural (PLN) é uma área da Inteligência Artificial que se dedica ao desenvolvimento de técnicas e algoritmos para a interpretação e manipulação de dados linguísticos em formato de texto. O PLN tem como objetivo principal capacitar os computadores a compreenderem e gerarem linguagem humana de forma semelhante aos seres humanos. No contexto deste projeto, o PLN desempenha um papel fundamental na identificação e extração de informações (Freitas, 2022) relevantes das receitas culinárias. Isso envolve tanto as receitas cadastradas pelos usuários quanto as receitas coletadas por meio do crawler, um programa de computador responsável pela mineração de dados na web.

Ao criar uma receita, o aplicativo utiliza técnicas de PLN para reconhecer automaticamente os ingredientes presentes no texto por meio de palavras-chave. Isso permite que os futuros usuários encontrem a receita cadastrada com base nos ingredientes específicos que desejam utilizar, facilitando a busca por opções adequadas às suas preferências e restrições alimentares. Além disso, o PLN também é aplicado na etapa de mineração de dados, onde o crawler coleta receitas disponíveis na internet. Nessa fase, o PLN é empregado para identificar e extrair os ingredientes das receitas encontradas, contribuindo para a expansão do banco de dados de receitas disponíveis no sistema. Com uma base de dados ampla e diversificada, os usuários terão acesso a uma variedade de opções culinárias para explorar e experimentar.

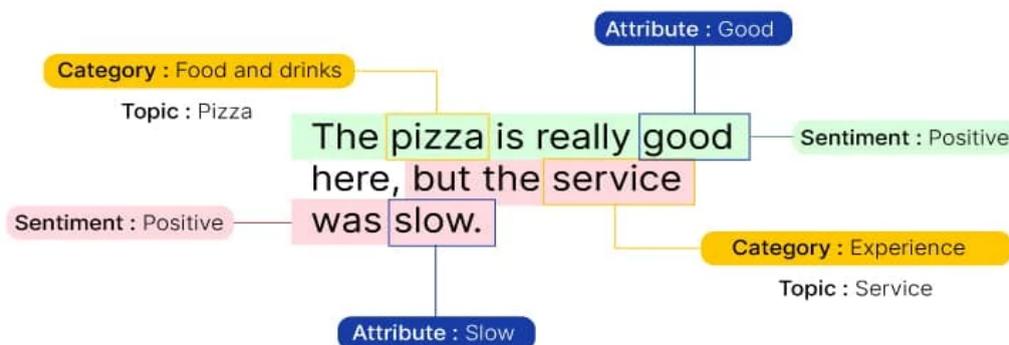


Figura 01 - Processamento de Linguagem Natural. Fonte: Bootcamp Uxdesign.

A imagem acima mostra uma frase que descreve a experiência em um restaurante de pizza, mencionando a qualidade da pizza e o serviço lento. Ao analisar essa frase, a PNL examina as palavras-chave e as associações emocionais evocadas, levando em consideração a satisfação com a pizza e possíveis frustrações causadas pelo serviço. Esses elementos têm impacto na percepção geral do restaurante. Além disso, a PNL também analisa a estrutura gramatical da frase, observando a ordem em que os aspectos são mencionados. Com base nesses insights, a PNL ajuda a compreender como interpretamos e avaliamos as informações ao pesquisar receitas culinárias e experiências gastronômicas.

Em suma, o uso do Processamento de Linguagem Natural neste projeto permite que o aplicativo reconheça e manipule a linguagem humana de forma inteligente, tornando mais fácil e eficiente a busca e a seleção de receitas com base nos ingredientes desejados. Essa abordagem amplia as possibilidades de escolha e proporciona uma experiência mais personalizada e adaptada às preferências dos usuários.

2.3 Mineração de Dados

A mineração de dados é o processo de descobrir padrões, tendências e informações relevantes em grandes conjuntos de dados, utilizando técnicas de análise e algoritmos de aprendizado de máquina (FERNANDES et al, 2019.). No contexto deste projeto, a mineração de dados é empregada para coletar e analisar receitas culinárias disponíveis na *internet*, enriquecendo a base de dados do aplicativo e oferecendo aos usuários uma variedade maior de opções para suas buscas.

Por meio da seguinte técnica de mineração de dados na *web*, conhecida também como *crawler*, foi possível realizar a extração de informações de sites de receitas culinárias. Essa técnica tem como função obter o *schema* de um site, ou seja, a estrutura de dados que compõe o site de forma organizada para os buscadores *web*, como o Google.



Figura 02 - Web Scraping. Fonte: Instituto Nacional de Estatística.

Na imagem acima (Figura 02), vemos os elementos essenciais desse processo. Um navegador web é utilizado para acessar o site desejado, enquanto uma requisição HTTP é feita para obter o código-fonte da página. Em seguida, é realizada a extração das informações relevantes, utilizando técnicas de seleção e filtragem de dados. Essas informações são então armazenadas, geralmente em formatos como planilhas ou bancos de dados, permitindo sua posterior análise e utilização. O web scraping é uma técnica

valiosa para obter dados de maneira automatizada e pode ser aplicada em diversas áreas. É uma ferramenta poderosa para obter insights valiosos a partir da vasta quantidade de informações disponíveis na web.

O *crawler* também foi utilizado para realizar o cadastro das receitas previamente mineradas dentro da plataforma de busca de receitas por ingredientes, o que tornou a integração com o algoritmo de PLN mais simples. Dessa forma, a mineração de dados permitiu que fosse possível reunir uma base de dados massiva e de maneira rápida, com receitas que são utilizadas no dia a dia da culinária brasileira.

3. Métodos

Uma vez que o presente trabalho se propõe a desenvolver uma aplicação e avaliar sua aderência, foram adotadas diversas abordagens metodológicas, incluindo pesquisa bibliográfica, pesquisa qualitativa e estudo de viabilidade com a aplicação de questionários e o uso de tecnologias específicas para o desenvolvimento do aplicativo e a mineração de dados.

A pesquisa bibliográfica foi conduzida principalmente na plataforma *Google Scholar*, que disponibiliza uma ampla variedade de artigos, livros e publicações acadêmicas. As palavras-chave empregadas foram "*recipes*", "*cook*", "*data mining*", "*crawler*" e "*natural language processing*". Os materiais mais relevantes foram selecionados com base na quantidade de palavras-chave presentes nas publicações.

A pesquisa qualitativa, por sua vez, envolveu a aplicação de um questionário eletrônico do tipo fechado, com o objetivo de compreender as experiências culinárias do público-alvo e obter *insights* para o desenvolvimento do aplicativo. O questionário foi compartilhado pelos membros da equipe em suas redes sociais e canais de comunicação, resultando em um total de 146 respostas.

O questionário não foi direcionado a um público-alvo específico, pois o objetivo era obter uma amostra mais representativa da população em geral. Com relação ao estudo de viabilidade, foi realizada uma análise dos dados primários coletados por meio do questionário, visando identificar possíveis falhas, sugestões de melhorias e ajustes necessários para atender às expectativas do público-alvo, bem como a aderência da proposta.

O desenvolvimento do aplicativo envolveu a utilização das seguintes tecnologias:

- **Design:** Para a criação dos protótipos e simulação dos produtos finais, a equipe optou por utilizar o Figma. Essa ferramenta foi escolhida devido à sua interatividade e capacidade de colaboração entre os membros da equipe. Além disso, a escolha do Figma foi influenciada pela familiaridade que a equipe já possuía com essa ferramenta;
- **Crawler:** Para realizar a coleta de dados e a manipulação dos elementos dos sites, foi utilizado o Node.js. Essa escolha se deu pela compatibilidade do Node.js com bibliotecas que permitem simular navegadores e manipular os

elementos dos sites. Essa flexibilidade proporcionada pelo Node.js permitiu que a equipe implementasse o crawler de forma eficiente;

- **Back-end (e PLN):** O back-end do aplicativo Chelf foi desenvolvido utilizando a linguagem C# e o framework .NET Core 6.0. Essa escolha foi feita levando em consideração o alto desempenho e a compatibilidade entre plataformas oferecidas pelo .NET Core. Além disso, o C# é uma linguagem poderosa que permite o desenvolvimento de aplicações robustas e de alta complexidade. A utilização de técnicas de Processamento de Linguagem Natural (PLN) no back-end possibilita a análise e a compreensão das informações textuais relacionadas às receitas culinárias, o que é essencial para o aplicativo de busca de receitas;
- **Front-end:** No front-end, foram utilizadas as tecnologias JavaScript, React e Next.js. O JavaScript é uma linguagem amplamente utilizada para o desenvolvimento web e permite a criação de interfaces interativas. O React é um framework JavaScript que facilita a criação de componentes visuais reutilizáveis, proporcionando um alto desempenho e uma melhor organização do código. O Next.js é uma biblioteca que complementa o React, oferecendo recursos adicionais, como renderização do lado do servidor, otimização de desempenho e facilidade de desenvolvimento. Além disso, a linguagem TypeScript foi empregada para evitar problemas comuns do JavaScript, fornecendo recursos de tipagem estática e auxiliando na detecção de erros durante o desenvolvimento;
- **Banco de dados:** Foram utilizados dois tipos de bancos de dados. Para o PWA (Progressive Web App), foi escolhido o Microsoft SQL Server. Esse banco de dados oferece eficiência e suporte oficial para integração em servidores de hospedagem, como o Azure. Essa escolha proporciona uma base de dados robusta e confiável para o armazenamento das informações do aplicativo. Para o crawler, foi utilizado o MongoDB. O MongoDB é um banco de dados NoSQL que se destaca pela sua capacidade de lidar com grandes volumes de dados não estruturados. Essa escolha foi feita para otimizar a coleta e armazenamento das informações provenientes da web, facilitando a manipulação e a extração de dados relevantes para o aplicativo.

Ao longo do projeto foram adotadas técnicas e ferramentas de gerenciamento de projetos para organizar e gerenciar as tarefas. O método *Kanban*, implementado por meio da ferramenta Trello, foi utilizado para visualizar o fluxo das atividades e seus respectivos *status*, além de atribuir tarefas aos membros da equipe, evitar a duplicação de esforços e garantir a eficiência do projeto.

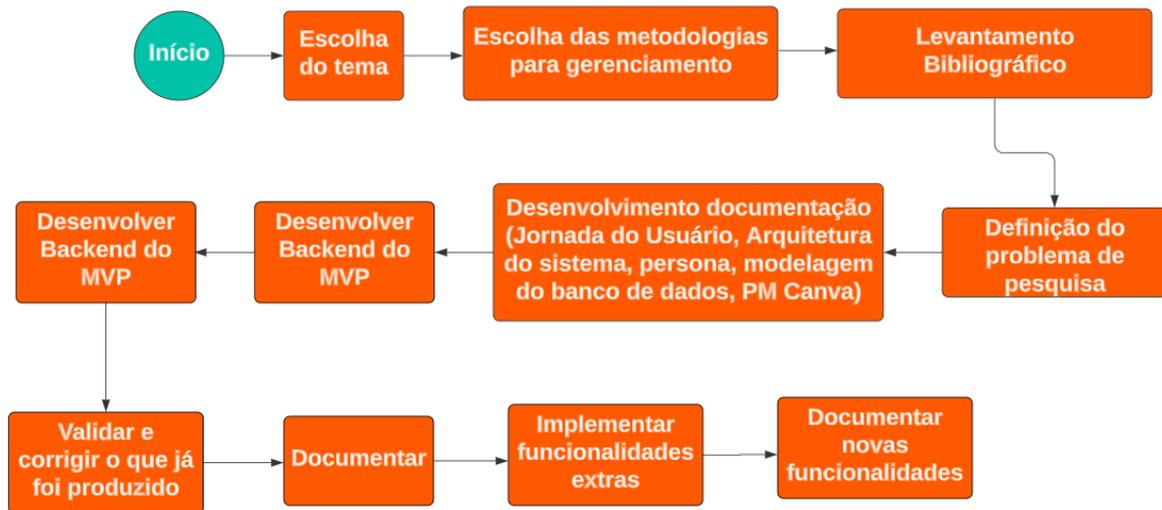


Figura 03 – Fluxograma método de pesquisa. Fonte: os autores

4. Desenvolvimento: O Sistema Chelf

O nome escolhido para o aplicativo é uma mistura de duas palavras: “*Chef*”, que significa “chefe de área de cozinha”. E a palavra “*self*” que traduzido do inglês pode significar: “Eu”, “auto”, “próprio”. Esse nome foi escolhido devido ao sistema proposto ser voltado para o ramo de culinária, e com o objetivo é permitir com que as pessoas tenham mais criatividade dentro da cozinha no preparo de suas refeições.

O sistema está disponível para acesso através do link a seguir: “chelf.app”, e é um sistema *Web*, com compatibilidade *mobile*, que pode ser instalado no dispositivo do usuário através do navegador. Pensando num sistema que fosse simples de utilizar e de rápido acesso, a página principal do sistema permite que o usuário realize a ação de procurar receitas com os ingredientes desejados. O fluxo do sistema para a página de busca de receitas é apresentado a seguir, na Figura 4.

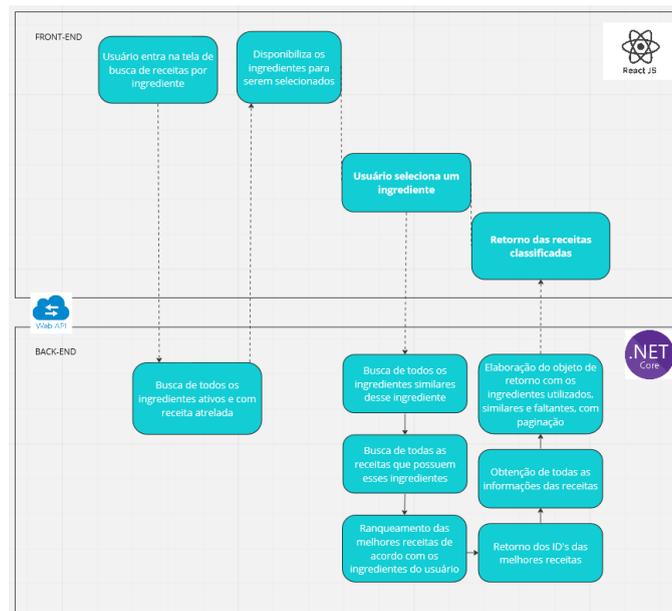


Figura 04 - Fluxo API da tela de busca através de ingredientes. Fonte: Os autores

4.1 Ingredientes relacionados e genéricos

Antes de explorar em detalhes as técnicas e padrões utilizados no desenvolvimento do sistema, é importante compreender o contexto em que ele está inserido. Em um aplicativo de receitas culinárias brasileiro, que abrange diversas culinárias regionais, existem ingredientes que podem ter nomes diferentes dependendo da localização. Por exemplo, a "Macaxeira" é conhecida em outras regiões do Brasil como "Aipim" e/ou "Mandioca".

Para lidar com essa variação de nomes de ingredientes, foi necessário pesquisar os principais alimentos que possuem essa característica de múltiplos nomes. Em seguida, esses alimentos foram armazenados e relacionados dentro do sistema, de modo que quando um usuário pesquisar por receitas contendo um ingrediente com o nome regional, o sistema será capaz de relacioná-lo com as receitas que possuem o nome mais conhecido. Por exemplo, ao pesquisar por "Aipim", o usuário encontrará receitas que contenham "Mandioca".

A questão linguística dos nomes de ingredientes também afeta os ingredientes genéricos, que são nomes de ingredientes que fazem parte do nome de outros ingredientes. Um exemplo é a "Farinha", que pode ser considerada isoladamente como "Farinha de trigo", mas também é usada para outros tipos de farinha, como "Farinha de rosca" e "Farinha de puba". Essa variação pode gerar ambiguidades na busca por receitas, pois um usuário que pesquisar por "Farinha" pode estar interessado em receitas que usem diferentes tipos de farinha.

Para solucionar esse desafio linguístico, foi adotada uma abordagem semelhante à mencionada anteriormente. Foi feito um levantamento de diversos ingredientes, e definiu-se que um ingrediente genérico é aquele que possui apenas uma palavra e outros ingredientes têm essa palavra como primeiro nome. Por exemplo, alimentos como "tempero" possuem uma única palavra e há outros alimentos que utilizam essa palavra como primeiro nome, como "tempero de frango" e "tempero de galinha".

Assim, os ingredientes genéricos foram identificados utilizando ferramentas de manipulação de dados, como o *Excel*. Após o levantamento, foi realizado um trabalho minucioso para validar a lógica aplicada, que se mostrou eficiente. Após a explicação do conceito de ingredientes relacionados e genéricos, é possível dar continuidade à leitura dos próximos tópicos.

4.2 Mineração de Dados e *Crawling*

Se por um lado a popularização das páginas web especializadas em receitas ajudaram a difundir a cultura do compartilhamento de receitas, por outro, a enorme quantidade de dados armazenados nesses repositórios torna a busca por uma receita que se adeque aos interesses dos usuários, uma tarefa árdua. Nesse contexto de dados massivos, técnicas de Mineração de Dados, como os Sistemas de Recomendação, são cada vez mais empregadas como ferramentas de suporte aos usuários na busca por informações úteis em seus contextos de aplicação (OLIVEIRA et al, 2019). O sistema desenvolvido utiliza técnicas de *Web Crawling* para coletar informações de receitas disponíveis na *web*. Nesta seção, serão abordadas as técnicas e padrões comuns nessas tecnologias, bem como a aplicação dessas tecnologias no contexto do trabalho.

4.2.1 Técnicas e Padrões Comuns em *Web Crawling*

Web Crawling é uma técnica amplamente utilizada para coletar informações de páginas *web* de maneira automatizada. Neste trabalho, a linguagem de desenvolvimento *NodeJS* foi utilizada em conjunto com a biblioteca *Puppeteer* como uma ferramenta de *Crawling* para navegar e extrair informações das páginas *web*. Entre os padrões comuns de *Crawling*, destacam-se:

- Exploração recursiva de *links*: o *crawler* segue os *links* encontrados nas páginas para coletar informações adicionais de outras páginas relacionadas;
- Extração de informações estruturadas: o *crawler* identifica padrões nos documentos HTML das páginas *web*, como seletores CSS ou XPath, para extrair informações específicas, como nome, categoria, ingredientes e tempo de cozimento das receitas;

- Estratégias de manobras para evitar bloqueios: o robô utiliza informações falsas, como falsos *headers-agents* e outras técnicas para driblar as defesas de DDOS e evitar ser barrado ao acessar o *site* várias vezes ao dia.

4.2.2 Armazenamento e Organização das Informações Coletadas

Após a extração das informações das páginas *web*, o *crawler* armazena os dados em um banco de dados *NoSQL MongoDB*, escolhido por questões de facilidade de implementação e consumo desses dados. No banco, são armazenadas todas as informações relevantes de uma receita, como nome, imagem, tempo de preparo, porções, ingredientes e medidas, modo de preparo, quantidade de *likes*, *link* para a receita, *tags* e categorias.

Outro *crawler* é responsável por consumir esse banco e cadastrar as receitas no *site* (*chelf.app*), que possui ligação com o sistema *backend* (C# .net), citado anteriormente, que processa as receitas utilizando PLN. Nesse novo processo, o *crawler* de cadastrar receitas no *site*, precisa acessar a aplicação através do *link* “*chelf.app*”, realizar o *login*, e navegar até o *link* de cadastro de receitas (Figura 5).

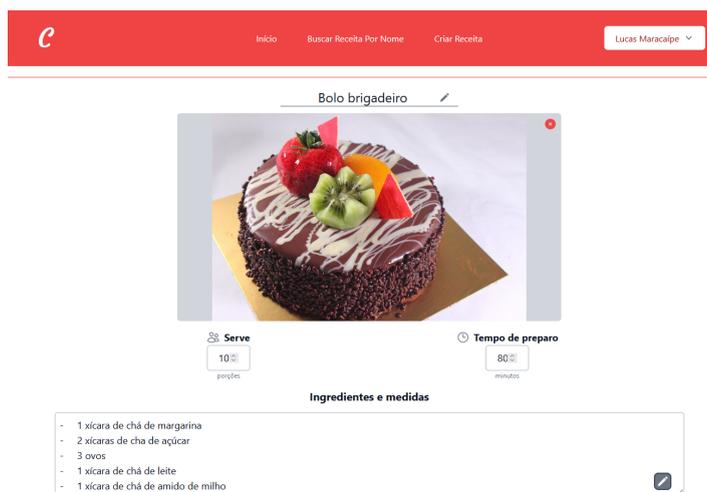


Figura 05 - Tela de cadastro de receitas. Fonte: os autores.

Dentro da aba de cadastro, o *crawler* busca as informações das receitas previamente mineradas contidas no banco de dados e começa a cadastrar uma a uma. O reconhecimento dos ingredientes contido nas receitas, através de PLN, é realizado durante o preenchimento desses dados nos campos de cadastro, via *Websocket*, onde futuramente o cadastrador seleciona os ingredientes sugeridos que foram reconhecidos pelo algoritmo de reconhecimento, como pode observar na Figura 6.

Ingredientes e medidas

- 1 xícara de chá de margarina
- 2 xícaras de chá de açúcar
- 3 ovos
- 1 xícara de chá de leite
- 1 xícara de chá de amido de milho



Ingredientes Selecionados 

Selecione os ingredientes e eles aparecerão aqui!

Sugestão de ingredientes com base no texto

Margarina Açúcar Ovo Leite Amido De Milho 

Figura 06 - Cadastro e reconhecimento de ingredientes utilizando PLN. Fonte: os autores.

4.3 Processamento de Linguagem Natural (PLN) na Identificação de Ingredientes

O sistema desenvolvido utiliza técnicas de *Web Crawling* em conjunto com o Processamento de Linguagem Natural (PLN) para a coleta e identificação dos ingredientes presentes nas receitas mineradas da *web*. Esta seção irá focar no emprego das técnicas de Processamento de Linguagem Natural e na forma como são aplicadas ao longo do projeto.

4.3.1 Abordagem do Processamento de Linguagem Natural (PLN) no Contexto do Trabalho

As técnicas de PLN são aplicadas para analisar e identificar ingredientes a partir do texto das receitas. Ao usar essas técnicas, o sistema consegue compreender e processar a linguagem humana de maneira eficiente e precisa. Isso permite que o sistema identifique os ingredientes no texto das receitas, mesmo que contenham erros de digitação.

No processo de cadastramento das receitas no *site* (*chelf.app*), o reconhecimento dos ingredientes, através de PLN, é realizado durante o preenchimento desses dados nos campos de cadastro, via *Websocket*. Após a sugestão dos ingredientes reconhecidos pelo algoritmo de PLN, o cadastrador seleciona os ingredientes adequados para serem inseridos na receita. Esse processo é ilustrado na Figura 7:

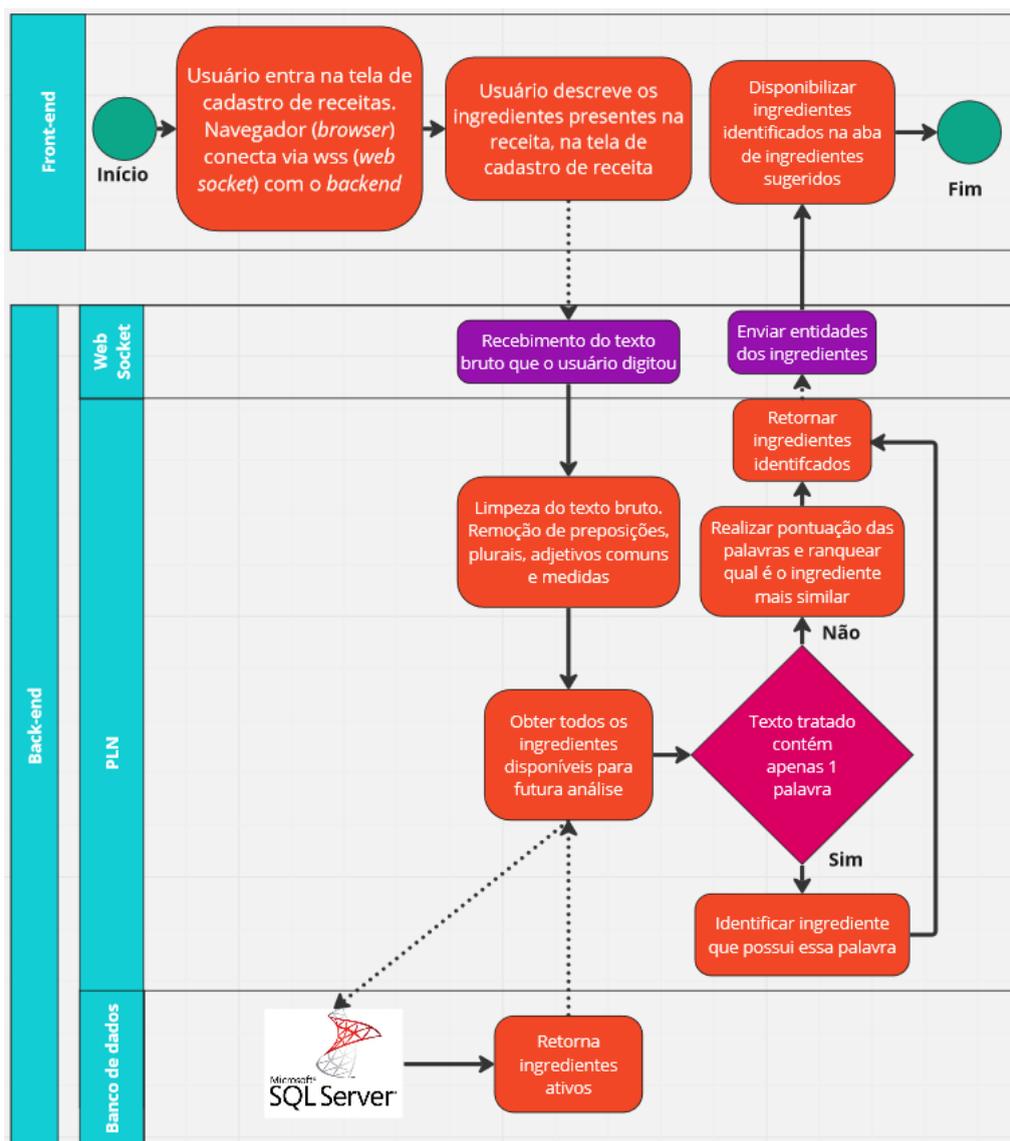


Figura 07 - Fluxo de identificação de ingredientes entre plataformas. Fonte: os autores

4.3.2 Contribuições do PLN para o Sistema

As técnicas de PLN desempenham um papel importante no sistema ao possibilitar a identificação e extração de ingredientes a partir do texto das receitas. A aplicação das técnicas de PLN ajudam a:

- Identificar os ingredientes mencionados nas receitas, independentemente de variações na escrita ou erros de digitação, como mostrado na figura 6.
- Ignorar palavras irrelevantes que não contribuem para o entendimento dos ingredientes, como preposições e artigos, através da remoção de *stopwords*.
- Comparar e classificar ingredientes com base em sua semelhança, utilizando abordagens baseadas em pontuação.

4.3.3 Pontuação e Classificação das Palavras na Identificação

Como citado anteriormente, o algoritmo de PLN está integrado com o texto que o usuário irá informar no campos de texto dos ingredientes, ele é integrado via *Web socket*, um tipo de conexão ativa, que o navegador cria com o *backend* do sistema para que toda vez que o usuário mudar o valor da caixa de texto, a mensagem ser reprocessado pelo algoritmo.

Assim que recebida a mensagem, o texto é tratado com as técnicas previamente citadas, permitindo assim que palavras sem valor não permaneçam no texto, ou seja, caso o usuário digite uma frase como “1 kg de farinha”, o numeral “1”, a medida “kg” e a preposição “de”, dessa frase serão removidos, tornando a frase em uma única palavra: “farinha”. Dessa forma, o sistema precisa identificar apenas uma palavra que, ao realizar o cruzamento dela com os ingredientes previamente cadastrados, que possuem essa mesma palavra, é reconhecido qual ingrediente o usuário quis dizer.

Para frases que mesmo após o tratamento possuem mais de uma palavra para serem reconhecidas, a complexidade aumenta drasticamente, pois os cenários de reconhecimento aumentam. Os cenários de identificação foram ilustrados na Figura 8:

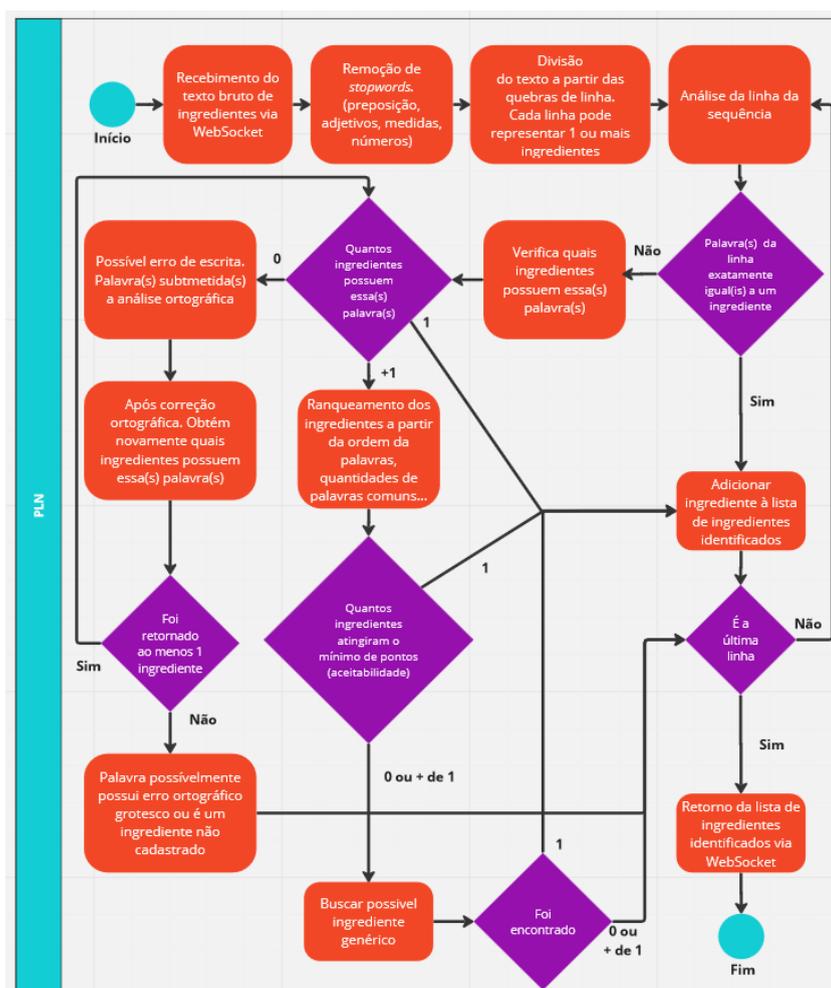


Figura 08 - Fluxo detalhado de identificação de ingredientes. Fonte: os autores

A combinação de técnicas de PLN e *Web Crawling* permite ao sistema oferecer uma solução eficiente e personalizada para os usuários que buscam informações sobre receitas e ingredientes na *web*.

5. Resultados e Discussão

A análise dos dados obtidos nas duas entrevistas, uma quali-quantitativa e outra com operação assistida seguida de coleta de *feedback*, que podem ser visualizadas no apêndice deste artigo, permitiu identificar padrões de comportamento e preferências dos usuários, bem como suas opiniões sobre a aplicação de Processamento de Linguagem Natural (PLN) e Mineração de Dados na criação de um aplicativo PWA de busca de receitas culinárias, com mais de 4000 receitas cadastradas.

Ao olhar sua geladeira e armários, você consegue idealizar quais receitas estão disponíveis ao seu alcance?

146 respostas

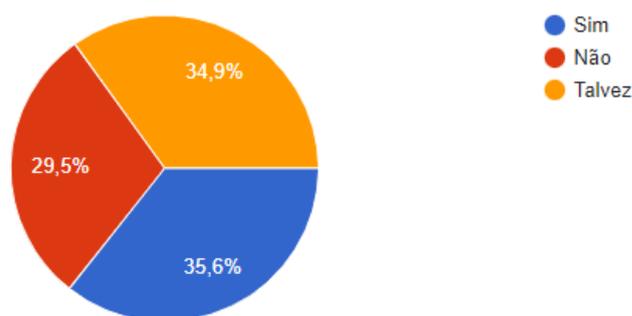


Figura 09 - Questionário. Fonte: os autores

Ao serem questionados com a pergunta “ao olhar sua geladeira e armários, você consegue idealizar quais receitas estão disponíveis ao seu alcance?” conforme demonstrado no gráfico acima, apenas 35,6% dos entrevistados responderam que sim, os outros 64,4% dos entrevistados responderam que não (29,5%) ou talvez (34,9%). Ainda nesse contexto, a maioria dos entrevistados na primeira entrevista (aproximadamente 76%) utiliza aplicativos de receitas todos os dias, demonstrando um alto interesse por esse tipo de plataforma. Além disso, a maioria (cerca de 70%) prepara suas próprias refeições, conforme pode ser visto no gráfico abaixo:

Quando você sente vontade de comer e não tem comida pronta disponível em casa, você...



146 respostas

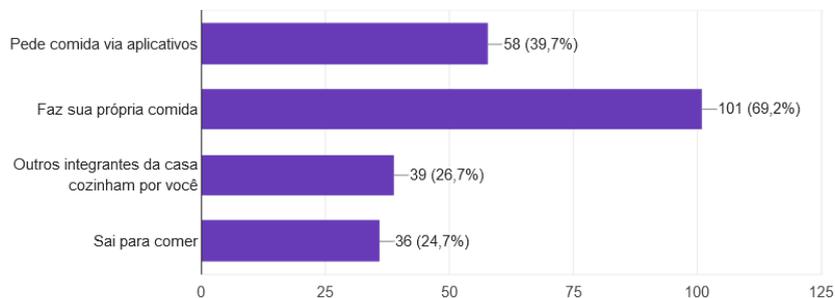


Figura 10 - Questionário. Fonte: os autores

A dimensão desse público também pode ser correlacionada com o gráfico acima, onde os entrevistados foram questionados com a pergunta “Quando você sente vontade de comer e não tem comida pronta disponível em casa, você...”: 69,2% dos entrevistados responderam que faz a sua própria comida e 26,7% que outros integrantes da casa cozinham. Portanto, de acordo com o questionário aplicado, há uma predominância do preparo caseiro de refeições, enquanto existe uma certa dificuldade em se idealizar receitas a partir de ingredientes disponíveis, confirmando a viabilidade e importância da funcionalidade de pesquisa de receitas à partir da seleção de ingredientes.

Na análise do *feedback* dos usuários sobre o aplicativo de receitas, os entrevistados da primeira entrevista expressaram satisfação com a facilidade de encontrar receitas (100%) e a rapidez do aplicativo (100%). Porém, sugeriram melhorias, como adicionar mais informações no menu, incluir mais imagens nas receitas e aumentar a variedade de pratos disponíveis.

Cerca de 50% dos usuários enfrentaram dificuldades ao cadastrar receitas devido à falta de alguns ingredientes e problemas na edição de textos. Os usuários também sugeriram melhorias no *layout*, adicionar uma seção de pratos favoritos e aumentar a variedade de receitas disponíveis. Para solucionar essas questões, é possível aplicar padrões e teorias de usabilidade, como as heurísticas de *Nielsen*, para aprimorar a interface e a experiência do usuário.

Na segunda entrevista, realizada com quatro usuários, foram identificadas sugestões para melhorar a experiência do usuário, como aumentar o tamanho da fonte, tornar a plataforma mais familiar e personalizada, e melhorar o design e adicionar mais cores. Esta entrevista contribuiu com aspectos que podem trazer uma melhor aceitação do aplicativo.

Os resultados mostram que o algoritmo de PLN foi capaz de reconhecer corretamente os ingredientes em 92% das receitas cadastradas. As dificuldades encontradas nos 8% restantes foram causadas por erros ortográficos grotescos e o uso de marcas como ingredientes. Para solucionar esse problema, é possível implementar um sistema de correção ortográfica automática e expandir a base de dados de ingredientes para incluir marcas populares.

Dessa forma, a análise das entrevistas oferece um panorama abrangente das expectativas e demandas dos usuários em relação ao aplicativo de busca de receitas culinárias. A aplicação de técnicas de PLN e Mineração de Dados, aliada a uma abordagem centrada no usuário e a adoção de padrões e teorias de usabilidade, pode contribuir significativamente para o sucesso e a popularidade da plataforma no mercado, ao mesmo tempo em que aprimora a experiência do usuário e fornece soluções cada vez mais personalizadas e eficientes na busca de receitas culinárias.

O chelf.app é uma ferramenta simples. Basta inserir os ingredientes que tem em casa e ele irá atualizar as receitas disponíveis. Com uma vasta coleção de receitas, poderá explorar e descobrir novos sabores de acordo com os ingredientes que possui. A seguir, estão disponíveis as imagens da tela inicial do aplicativo (figura 11), onde os usuários têm a opção de incluir os ingredientes disponíveis em casa. Na tela principal (figura 12), as receitas cadastradas são exibidas de maneira organizada, acompanhadas de fotos ilustrativas. A tela de detalhes da receita (figura 13) possui um *layout* visualmente atraente, fornecendo uma apresentação clara e separada dos ingredientes e do modo de preparo.

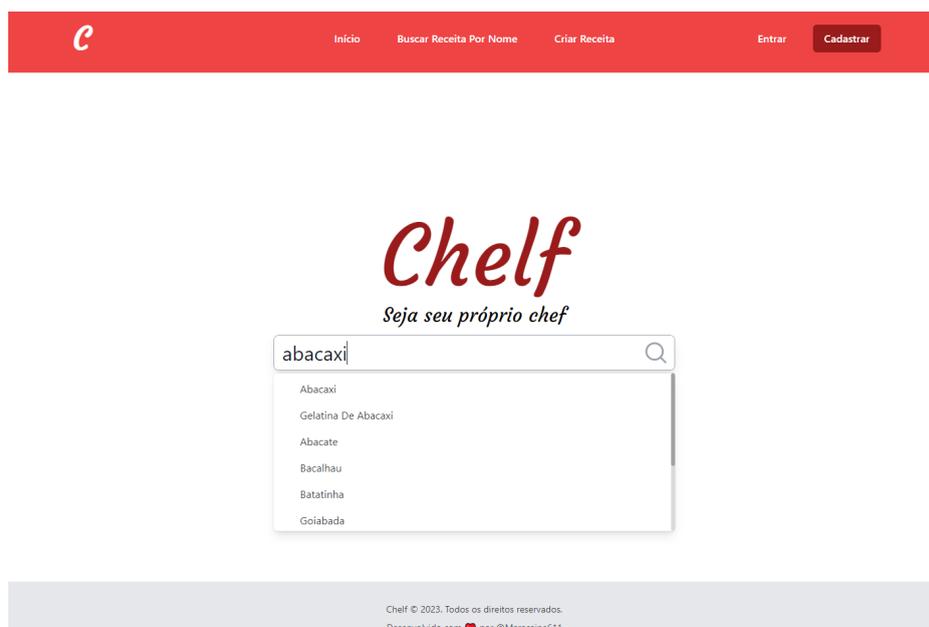


Figura 11 - Tela Inicial Chelf. Fonte: os autores

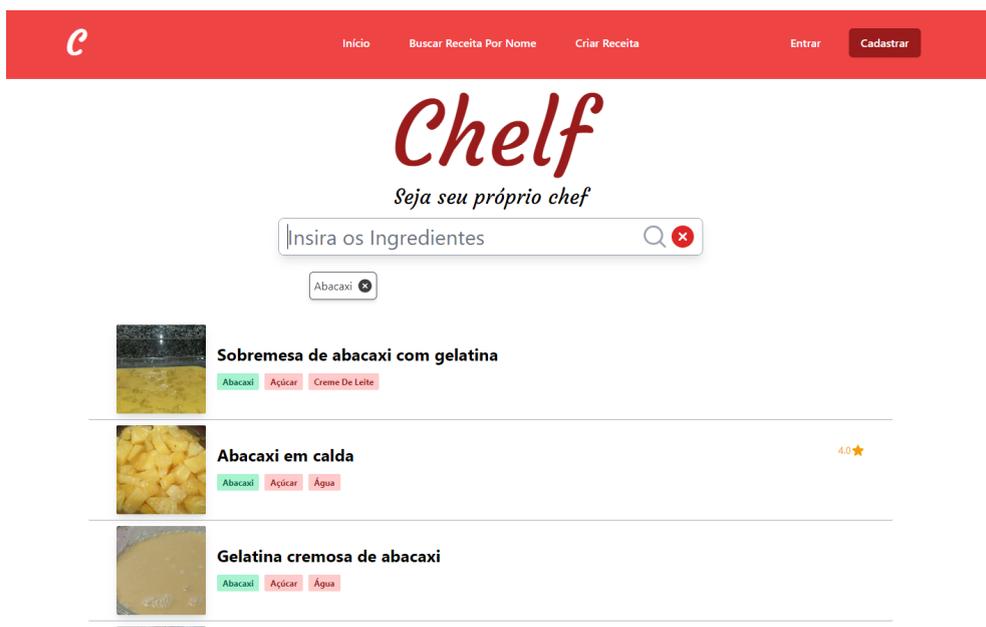


Figura 12 - Tela Principal Chelf. Fonte: os autores

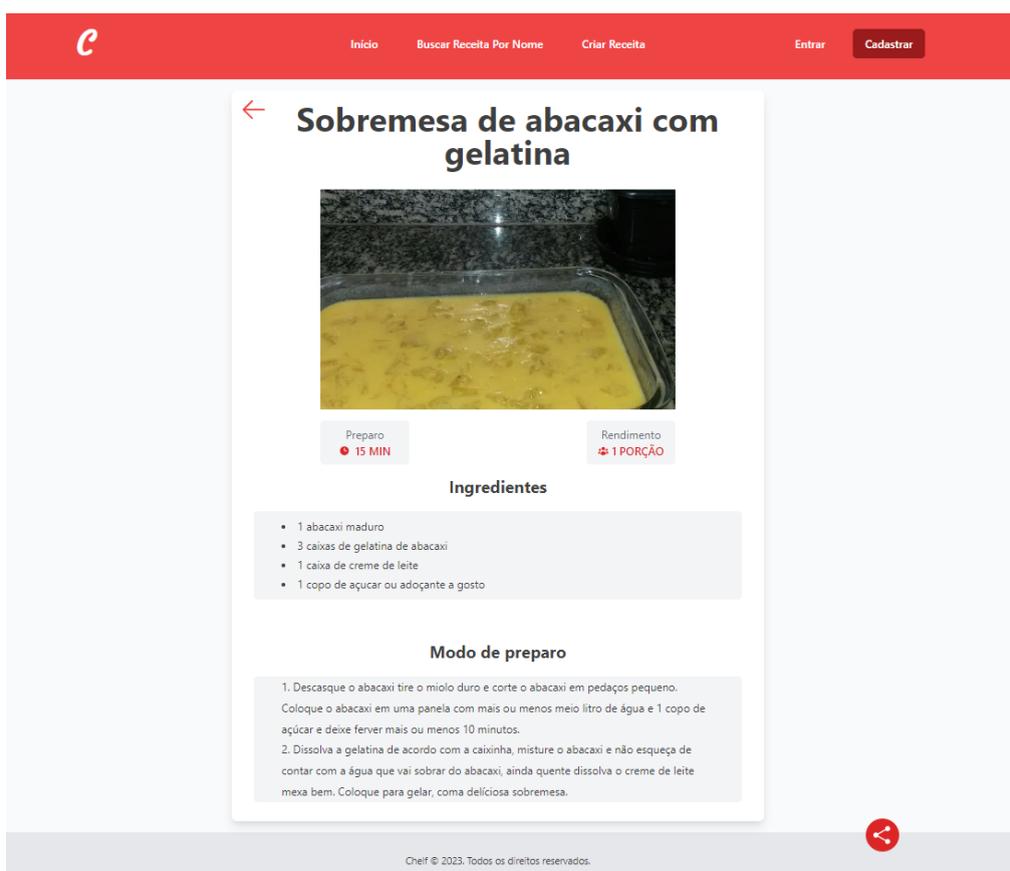


Figura 13 - Tela Detalhes da Receita Chelf. Fonte: os autores

6. Conclusão

Este estudo teve como objetivo geral desenvolver uma aplicação que utiliza técnicas de mineração de dados para fornecer sugestões e recomendações de receitas com base nos ingredientes informados pelo usuário. Para atingir esse objetivo, foram estabelecidos objetivos específicos:

1. Desenvolver um algoritmo de crawler para mineração de dados: Apesar dos desafios enfrentados com os sites minerados que identificaram o algoritmo como um bot de *DDOS*, foi possível desenvolver um algoritmo eficiente para coletar dados de diferentes fontes de receitas culinárias disponíveis na internet.
2. Desenvolver um algoritmo de Processamento de Linguagem Natural (PLN): O desenvolvimento do algoritmo de PLN foi bem-sucedido, embora tenha apresentado desafios na catalogação e relacionamento de ingredientes semelhantes. A capacidade de reconhecer automaticamente os ingredientes nas receitas foi essencial para permitir a busca e sugestão de receitas com base nos ingredientes informados pelos usuários.
3. Desenvolver um Progressive Web App (PWA): O objetivo de desenvolver um PWA foi alcançado, proporcionando aos usuários uma interface interativa e intuitiva. No entanto, a construção de uma interface compreensível e de fácil uso foi um desafio significativo, ressaltando a importância de considerar as necessidades e preferências dos usuários durante o processo de desenvolvimento.

Considerando os resultados obtidos, a aplicação de técnicas de Processamento de Linguagem Natural (PLN) e mineração de dados demonstrou potencial para proporcionar uma experiência satisfatória aos usuários na busca de receitas culinárias com base nos ingredientes. Os desafios enfrentados, como a identificação incorreta do algoritmo de crawler e a catalogação de ingredientes, mostraram a necessidade de implementar melhorias sugeridas, como sistemas de correção ortográfica automática e expansão da base de dados de ingredientes.

As conclusões obtidas orientam o aprimoramento futuro do aplicativo, visando sua utilidade e atratividade para os usuários. Sugere-se explorar mais a integração de PLN e mineração de dados, proporcionando soluções personalizadas e eficientes na busca de receitas culinárias. Para trabalhos futuros, é recomendado realizar estudos com mais usuários, validar a aplicação e coletar feedback. Além disso, investigar teorias de usabilidade no design da plataforma, otimizando a experiência do usuário e garantindo sua satisfação. Assim, o aplicativo poderá oferecer uma ferramenta mais eficaz na busca por receitas culinárias.

Em suma, este estudo demonstrou que a aplicação de técnicas de Processamento de Linguagem Natural (PLN) e mineração de dados oferece resultados satisfatórios na busca de receitas culinárias com base nos ingredientes. A abordagem adotada apresenta potencial para fornecer uma experiência atraente e satisfatória aos usuários, desde que as sugestões de melhorias e as demandas do público-alvo sejam consideradas no desenvolvimento de futuras versões do aplicativo.

7. Referências

- CAMPOS, Bruna Tafuri Lobato et al. Obesidade infantil na atualidade: fatores de risco e complicações futuras. *Brazilian Journal of Health Review*, v. 6, n. 2, p. 5838-5845, 2023.
- CAMPOLINO, Luana Moura. Análise integrativa do perfil multi omico dos canceres de mama e prostata. 2022. Universidade Estadual Paulista (Unesp), 2022. Disponível em: <http://hdl.handle.net/11449/238256>
- Brasil Ministério da Saúde (MS). Guia alimentar para a população brasileira. Brasília: MS; 2014. [Acessado 14 agosto 2022]. Acesso em https://bvsmms.saude.gov.br/bvs/publicacoes/guia_alimentar_populacao_brasileira_2e_d.pdf
- MAZZONETTO, Ana Cláudia, Dean, Moira e Fiates, Giovanna Medeiros Rataichesk. Percepções de indivíduos sobre o ato de cozinhar no ambiente doméstico: revisão integrativa de estudos qualitativos. *Ciência & Saúde Coletiva* [online]. v. 25, n. 11 [Acessado 14 Agosto 2022], pp. 4559-4571. Acesso em: <https://doi.org/10.1590/1413-812320202511.01352019>
- PORCH. Cooking Nightmares, A generational look capabilities in the kitchen. [Acessado 05 setembro 2022]. Acesso em <https://porch.com/resource/cooking-nightmares>
- ARAGÃO, Paulo André Pimenta Scrum XPerience: um protótipo de jogo voltado para o ensino de metodologias ágeis. Universidade Estadual Paulista (Unesp), 2022. Acesso em: <<http://hdl.handle.net/11449/216173>>
- FREITAS FILHO, Nelson Silva de. [2022] Aprendizado de máquina e processamento de linguagem natural aplicado na predição de marcas a partir de dados de marketplaces. Acesso em <https://repositorio.ufrn.br/handle/123456789/49080>
- GANDHI, S. (2022, April 25). Quantitative UX Research Methods - Bootcamp. Medium; Bootcamp. [Acessado 06 julho 2023]. Acesso em: <<https://bootcamp.uxdesign.cc/quantitative-ux-research-methods-70cebd7af63>>
- INE. O uso do Web Scraping nas estatísticas oficiais. [s.l: s.n.]. Disponível em: <https://www.ine.pt/ngt_server/attachfileu.jsp?look_parentBoui=292691736&att_display=n&att_download=y>. Acesso em: 6 jul. 2023.
- FERNANDES, Fernando Timoteo; CHIAVEGATTO FILHO, Alexandre Dias Porto. Perspectivas do uso de mineração de dados e aprendizado de máquina em saúde e segurança no trabalho. *Revista Brasileira de Saúde Ocupacional*, v. 44, 2019. <https://www.scielo.br/j/rbso/a/NgxW5qxzQWhcD4KrTHLxxGG/abstract/?lang=pt>
- OLIVEIRA, Emilia; BRITTO, Larissa; PACÍFICO, Luciano; LUDERMIR, Teresa. Recomendação e Geração de Receitas Baseada na Substituição de Ingredientes. In: ENCONTRO NACIONAL DE INTELIGÊNCIA ARTIFICIAL E COMPUTACIONAL (ENIAC), 16. , 2019, Salvador. Anais [...]. Porto Alegre: Sociedade Brasileira de Computação, 2019. p. 238-249. ISSN 2763-9061. DOI: <https://doi.org/10.5753/eniac.2019.9287>.

NOGUEIRA, T. da S. SIQUEIRA, K.B. GOLIATT, P.V.Z.C.
<https://www.alice.cnptia.embrapa.br/handle/doc/1136712>

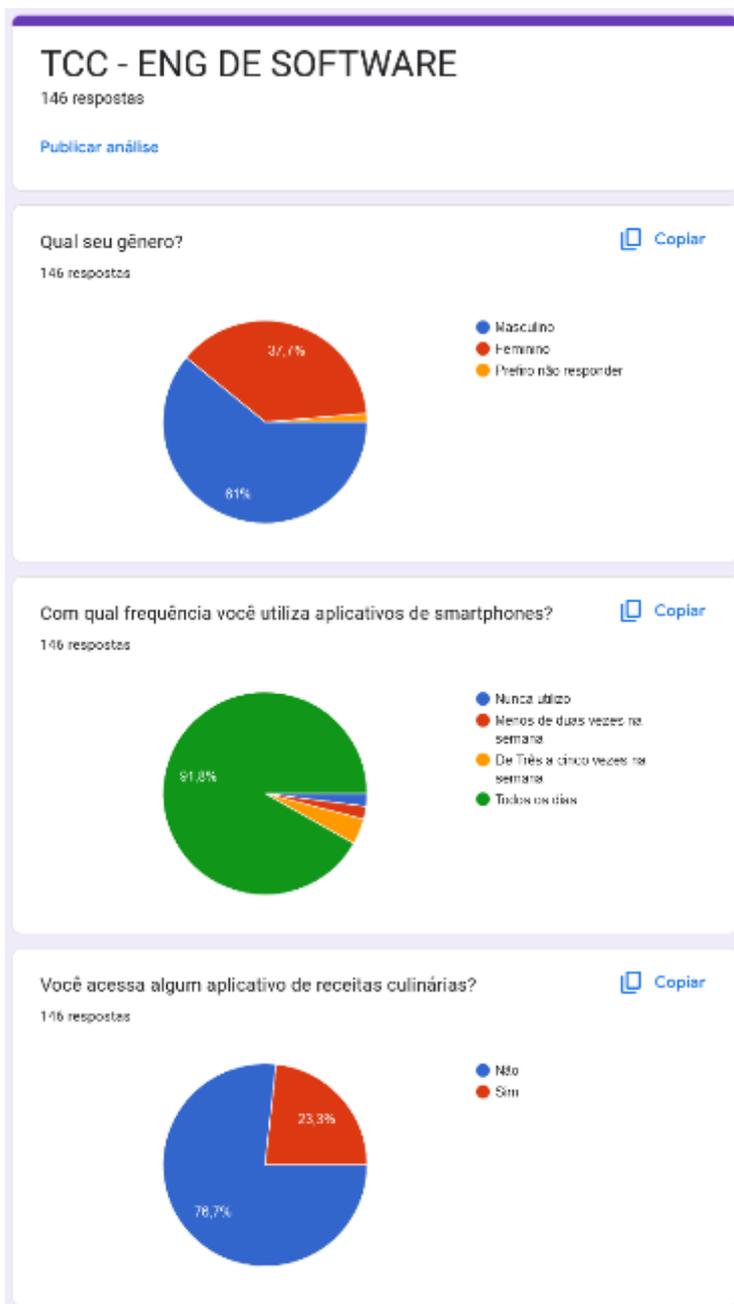
Luciano D. S. Pacífico, Larissa F. S. Britto, Teresa B. Ludermir
<https://sol.sbc.org.br/index.php/semish/article/view/15821>

BAGUETE, Webedia: R\$ 49 mi por TudoGostoso, 2015. [Acessado 05 setembro 2022].
Acesso em:
<https://www.baguete.com.br/noticias/21/08/2015/webedia-r-49-mi-por-tudogostoso>

Thallys da Silva Nogueira, Kenya Beatriz Siqueira, Priscila Vanessa Zabala Capriles
Goliatt <https://sol.sbc.org.br/index.php/sbiagro/article/view/18389>

Gartner. (2021). “Definition of Big Data”, Acesso em
<https://www.gartner.com/en/information-technology/glossary/big-data>.

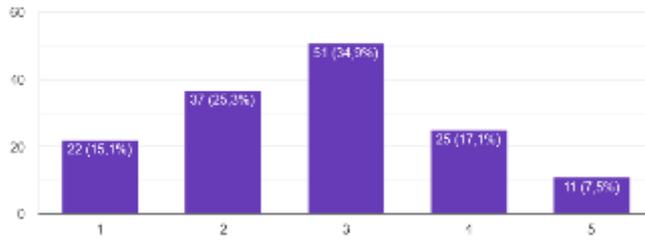
8. Apêndice



Copiar

Classifique seu nível de experiência na cozinha.

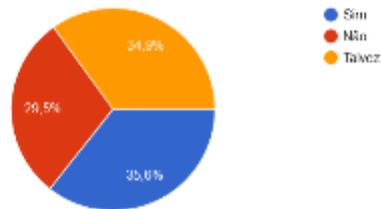
146 respostas



Copiar

Ao olhar sua geladeira e armários, você consegue idealizar quais receitas estão disponíveis ao seu alcance?

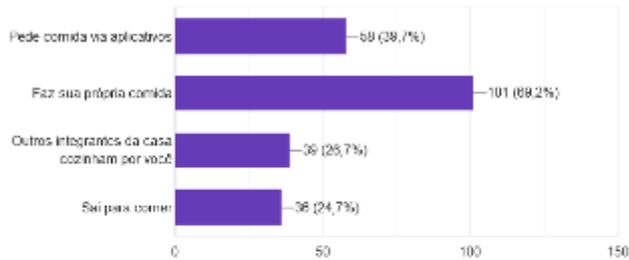
146 respostas



Copiar

Quando você sente vontade de comer e não tem comida pronta disponível em casa, você...

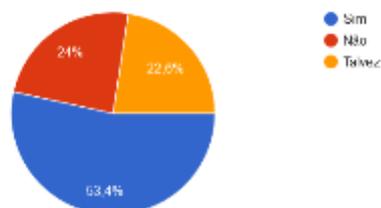
146 respostas



Copiar

Você se considera uma pessoa com tempo disponível para cozinhar pra si mesmo? Levando em conta que uma refeição rápida leva em torno de 15 minutos

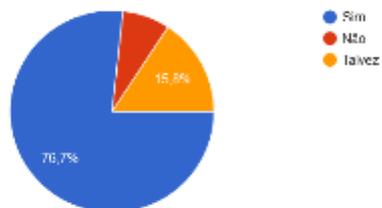
146 respostas



Caso você seja uma pessoa inexperiente na cozinha, você se arriscaria a realizar receitas se soubesse quais estão ao seu alcance com os ingredientes que você possui em casa?

[Copiar](#)

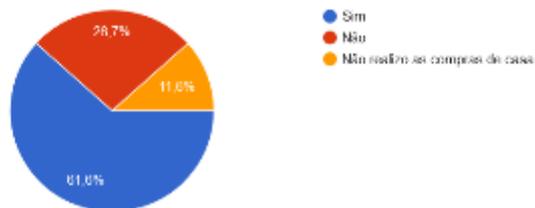
146 respostas



Quando você vai ao mercado comprar seus mantimentos, já compra ingredientes culinários pensando nas receitas que irá realizar?

[Copiar](#)

146 respostas



Este conteúdo não foi criado nem aprovado pelo Google. [Denunciar abuso](#) - [Termos de Serviço](#) - [Política de Privacidade](#)

Google Formulários

Entrevista pós operação assistida - Chelf

4 respostas

[Publicar análise](#)

Nome do participante

4 respostas

Leila

Denise

Luiz Fernando

Cristiane

1. Do contato que você teve com outros sites de receita, você sente falta de alguma funcionalidade?

4 respostas

Não

No menu, sentiu falta de mais informação. Sobre o perfil, contato, etc
Sentiu falta de mais imagens nas receitas

Não encontrou alguns pratos que esperava encontrar

Ela não achou tão intuitivo.

2. Do que você gostou?

4 respostas

Estar com a receita na mão - Clicou, achou a receita

Rapidez

Ter o chelf como um caderno de anotações para as próprias receitas
Ter a possibilidade de encontrar receitas com os ingredientes que eu gosto

Ela pode cadastrar as próprias receitas

3. Do que você não gostou?

4 respostas

Formular a receita, porque não achei alguns ingredientes, e me atrapalhei em alguns passos

Erros na hora da edição

Parte de cadastro. Achou complicada

O tamanho da letra é pequeno. Sentiu falta de poder aumentar a letra

3. Você dificuldade em realizar alguma ação?

4 respostas

Cadastrar a receita

-

Adicionar receitas, sentiu que tinha que entender sobre a receita e ter começar do zero a parte da receita

Ela não gostou do menu.

4. O que você melhoraria na forma que as receitas são pesquisadas e listadas?

4 respostas

Nada

O layout estava muito grande

Colocaria uma noção de pratos favoritos, deixando as receitas mais a ver com o contexto e perfil

Sentiu falta de variedade de receitas. Realizou inúmeras buscas por nome e n teve nenhum resultado.

5. O que você melhoraria na forma que as receitas cadastradas?

4 respostas

A identificação dos ingredientes. Talvez falte ingredientes, então falta a identificação correta

Nada

Parte de ingredientes

Sentiu falta de informar que existem variações para a receita que cadastrou

6. Teria algum comentário, feedback ou sugestão a mais que você gostaria de dar?

4 respostas

O usuário precisa ser mais familiarizado.

Mais opções para acessar as páginas. Parece que está faltando alguma coisa. Menu muito pequeno.

Deixar o Chelf mais personalizado

Ex: Qual ingrediente mais gosta

Qual tipo de comida gosta mais

Para que as sugestões sejam mais de acordo com o perfil.

Melhorar o design, colocar mais cores. Achou o site meio monocromático

Este conteúdo não foi criado nem aprovado pelo Google. [Denunciar abuso](#) - [Termos de Serviço](#) - [Política de Privacidade](#)

Google Formulários