

CENTRO FEDERAL DE EDUCAÇÃO TECNOLÓGICA DE
MINAS GERAIS – CAMPUS V

Sistema de Recomendação de Receitas Saudáveis - Tudo Saudável

Ana Beatriz Silveira Assis

Larissa Costa Vieira

Thaís Rita Silva Rosa

Divinópolis - MG

2015

CENTRO FEDERAL DE EDUCAÇÃO TECNOLÓGICA DE
MINAS GERAIS – CAMPUS V

Sistema de Recomendação de Receitas Saudáveis - Tudo Saudável

Ana Beatriz Silveira Assis

Larissa Costa Vieira

Thaís Rita Silva Rosa

Orientador: Willyan Michel Ferreira

Trabalho de Conclusão de Curso
apresentado ao Curso Técnico em
Informática do Centro Federal de
Educação Tecnológica de Minas
Gerais – Campus V como requisito
parcial para a obtenção do título de
Técnico em Informática.

Divinópolis

2015

CENTRO FEDERAL DE EDUCAÇÃO TECNOLÓGICA DE
MINAS GERAIS – CAMPUS V

Trabalho de Conclusão de Curso julgado adequado para obtenção do título de Técnico em Informática e aprovado pela banca composta pelos seguintes professores.

Prof. Willyan Michel Ferreira - CEFET-MG (Orientador)

Prof. Thiago Magela Rodrigues Dias - CEFET-MG

Prof. Vinícius Larêdo Henriques Duarte - CEFET-MG

Prof. Luís Augusto Mattos Mendes
Coordenador do Curso Técnico em Informática

Data de aprovação: Divinópolis, 25 de novembro de 2015.

RESUMO

Sistemas de recomendação são mecanismos que vem ganhando força na área de Recuperação da Informação (RI) aplicada à *Web*, a qual envolve a aplicação de métodos computacionais no tratamento da informação para avaliar em que medida a Ciência da Computação contribui para o avanço da Ciência da Informação. Nesse âmbito, encontram-se sistemas voltados a diversos assuntos, tais como filmes, livros, músicas e todo tipo de comércio *online*. Contudo, considerando-se a crescente preocupação dos brasileiros com a saúde e principalmente com seus hábitos alimentares, nota-se a falta de um sistema que recomende opções para tornar receitas culinárias, assunto de diversos sites especializados, em mais saudáveis. É nesse contexto que surge o Sistema de Recomendação de Receitas Saudáveis - Tudo Saudável, um sistema que possui por principal intuito estimular a mudança dos hábitos alimentares dos brasileiros. Através da coleta, tratamento e desambiguação dos ingredientes mais presentes nas receitas culinárias encontradas no site Tudo Gostoso, foram elaboradas recomendações, a partir do valor calórico encontrado de cada ingrediente, proporcionando assim uma opção mais saudável ao usuário que procura manter uma alimentação balanceada, dentro dos limites calóricos indicados pelos nutricionistas.

Palavras-chaves: Recuperação da Informação; Receitas culinárias; Alimentação Saudável

SUMÁRIO

1. Introdução	8
1.1. Definição do Sistema.....	9
1.2. Definição do Escopo.....	9
2. Referencial Teórico.....	9
2.1. Alimentação e Saúde	10
2.2. Sistemas de Recomendação <i>Web</i>	11
2.3. Métricas para Análise de Dados	12
2.4. Coletor de Dados	14
2.5. Bancos de Dados Não Relacionais	15
3. Projeto do Sistema	17
3.1. Definição das Funcionalidades	17
3.2. Coleta, análise e desambiguação dos ingredientes das Receitas	18
3.3. Pesquisa de níveis calóricos dos ingredientes	20
3.4. Formulação das recomendações	20
3.5. Estruturação do banco de dados	20
3.6. Desenvolvimento da Interface <i>Web</i>	21
4. Projeto Conceitual	21
4.1. Diagrama de Caso de Uso	22
4.2. Documentação do Ator	22
4.1.1.1. Usuário	22
4.3. Diagrama de Implantação	23
4.4. Diagrama de Classe	24
5. Projeto Físico.....	24
5.1. DER – Diagrama de Entidade e Relacionamento.....	24
5.2. DTR - Diagrama de Tabelas Relacionais	26
6. Resultados Finalizados.....	26
7. Considerações Finais	32

8. Cronograma	33
9. Referências	34
10. Apêndices.....	35
10.1. Apêndice A: Script de Coleta	35
10.2. Apêndice B: Script de Identificação	38
10.3. Apêndice C: Descrição Caso de Uso	41

1. Introdução

De acordo com a Vigitel (2014), a preocupação dos brasileiros com a saúde e as mudanças de hábitos alimentares tem sido crescente. Nesse contexto, viu-se a oportunidade para a criação de um novo sistema, que possibilitasse a conveniente união das pesquisas acerca do tema já perpetradas até o momento, com uma ferramenta de utilidade prática e acessível. Sendo assim, o Sistema de Recomendação de Receitas Saudáveis - Tudo Saudável, consiste em uma forma de proporcionar aos usuários de sites culinários uma maneira de buscar receitas culinárias menos calóricas.

O Sistema aqui explicitado tem por proposta apresentar substituições de ingredientes viáveis a partir dos ingredientes usados nas receitas de base extraídas do site Tudo Gostoso. Essa base foi estruturada através de um coletor de dados que, segundo Liddy (2000), é um processo de análise natural de ocorrência de texto com a finalidade de descobrir e capturar informações semânticas. Com as correlações dos ingredientes e seus respectivos níveis calóricos, são indicadas ao usuário opções que possam substituí-los, tornando, assim, o alimento final preparado mais saudável, pela recomendação de dada substituição.

A seguir serão apresentados detalhes do Sistema de Recomendação de Receitas Saudáveis - Tudo Saudável (Figura 1), bem como das metodologias empregadas em seu processo de criação. Serão definidos então, o escopo do projeto, os objetivos do sistema e as funcionalidades abrangidas por ele, além das etapas do desenvolvimento e as ferramentas utilizadas pela equipe. Por fim, os diagramas e resultados gerados pelo sistema serão apontados de maneira objetiva e sucinta.



FIGURA 1: Logo do Sistema de Recomendação de Receitas Saudáveis - Tudo Saudável

1.1. Definição do Sistema

O Sistema de Recomendação de Receitas Saudáveis - Tudo Saudável, é uma ferramenta que torna receitas culinárias menos calóricas através de recomendações de substituições de ingredientes. Essas recomendações são orientadas pelo nível calórico dos ingredientes encontrados na base de dados coletada do site Tudo Gostoso. Portanto, através de uma base de ingredientes, é indicada ao usuário uma alternativa menos calórica para a substituição dos ingredientes solicitados, acarretando em um decréscimo de calorias nas receitas finais.

O sistema é voltado - de forma enfática, mas não exclusiva - para usuários de sites culinários que tenham interesse em hábitos alimentares mais saudáveis. Visto que foi implementado para plataforma *Web* e que foi projetado como possível apoio para um site de receitas gastronômicas, a amplitude do público alvo é extremamente diversificada.

1.2. Definição do Escopo

Através da análise dos níveis calóricos dos ingredientes recolhidos do Tudo Gostoso, o sistema se propõe a recomendar ao usuário uma alternativa menos calórica a dado ingrediente de uma receita. Essa recomendação torna a receita menos calórica sem alterar significativamente suas propriedades, muito embora o sistema não tenha entre seus objetivos principais garantir a aprovação do usuário quanto ao resultado final da receita escolhida. Para auxiliar o usuário em tal âmbito, é disponibilizado um método de avaliação das recomendações. Nesta avaliação, o usuário que assim desejar, pode comentar e classificar, votando em botões de “*like*” ou “*dislike*”, a substituição recomendada. Essa votação gera um ranking pontuado em taxa de aprovação e preferência, o que auxilia o usuário a encontrar a melhor opção entre as receitas disponibilizadas a ele. O projeto foi desenvolvido para o ambiente *Web*, sendo um site formulado a partir da ferramenta *Bootstrap*, e, a priori, não oferece suporte para receitas de outros sites que não o Tudo Gostoso.

2. Referencial Teórico

Nos tópicos seguintes serão explicitados os referenciais teóricos do Sistema de Recomendação de Receitas Saudáveis - Tudo Saudável. Tais pontos foram abordados, de forma direta ou indireta, no decorrer do desenvolvimento do projeto.

Assim, a fim de fornecerem um suporte mais completo para o projeto, esses temas foram pesquisados de forma mais aprofundada e expostos a seguir.

2.1. Alimentação e Saúde

A busca por um estilo de vida mais saudável tem sido cada vez mais tema de pesquisas e projetos. A prática de exercícios aliados a novos hábitos alimentares tem sido vertente crucial nesse sentido. Os resultados de uma alimentação equilibrada são refletidos na saúde, podendo-se observar diminuição de doenças cujas principais causas estão na má alimentação, como a obesidade e a hipertensão arterial.

A obesidade é uma condição que aumenta o risco de morbidade para as principais doenças crônicas: hipertensão, dislipidemia, diabetes, doença coronariana, alguns tipos de câncer e colecistite e, embora não se conheça uma estratégia adequada de cura, sua prevenção e tratamento apresentam-se como um dos grandes desafios deste século (TAUBES, 1998).

Parece, então, haver boas razões para encorajar mudanças alimentares aliados a exercícios físicos e perda de peso, mesmo entre pessoas com sobrepeso, promovendo a sua longevidade e qualidade de vida (WILLETT, 1997; STEVENS, 1998). Segundo Coitinho (1999), manter um peso corporal adequado e não ganhar peso durante a vida adulta parece associar-se a menor mortalidade e maior bem-estar. Uma alimentação equilibrada e a manutenção do peso ideal entram aí como chave para a prevenção das doenças crônicas não transmissíveis e promoção de melhor qualidade de vida (WILLETT, 1997).

Portanto, o estabelecimento de dietas saudáveis deve contemplar como prioridade a prevenção do ganho de peso. Para adultos considera-se como peso saudável, o peso relativo, avaliado pelo índice de massa corporal ($IMC = \text{peso em kg} / \text{altura}^2 \text{ em m}$) de até 24,9 (Organização Mundial da Saúde, 1998). Segundo Sichieri (1999), para crianças e adolescentes sugere-se também a utilização do IMC, contudo os pontos de corte adequados são ainda objeto de discussão.

Com esse aumento da preocupação acerca da obesidade e das doenças associadas ao peso inadequado, no Brasil, há que se combinarem orientações para a redução das deficiências nutricionais, ainda presentes, e da prevenção ao ganho de peso. Uma proposta eficiente de alimentação saudável há de propor dietas que estejam ao alcance da sociedade como um todo, e que tenham um impacto sobre os mais importantes fatores relacionados às várias doenças. Aumentar o consumo de frutas e verduras e estimular o consumo de arroz e feijão são exemplos de proposições que preenchem estes requisitos.

Do ponto de vista da alimentação saudável, sugere-se, então, que as recomendações devem basear-se em alimentos mais do que em nutrientes. Assim, a Organização Mundial de Saúde, em publicação recente (Organização Mundial de Saúde, 1998), sugere o estabelecimento de metas realísticas de consumo de alimentos específicos, sendo estes alimentos identificados em função dos nutrientes que se pretendam abranger e de seus baixos níveis calóricos.

No estabelecimento das recomendações para a população brasileira são consideradas relevantes as intervenções referentes à prevenção da obesidade, das doenças cardiovasculares, câncer, diabetes tipo 2 e osteoporose. Fomentar atividades de informação ao consumidor em meios acessíveis e estabelecer especificação para rotulagem pelo Ministério da Saúde também seriam metas importantes dentro desta ótica.

2.2. Sistemas de Recomendação Web

Desde sua criação, a *World Wide Web* sofre constantes mudanças de conteúdo e abrangência. A cada nova funcionalidade oferecida pelas aplicações *Web*, uma nova função é designada ao usuário, cujo papel passa de consumidor para produtor de conhecimento. Essa colaboração é inevitável para a construção da chamada “*Web Social*”, que engloba sites e *softwares* projetados a fim de promover a interação social. Sob essa perspectiva, os sistemas de recomendação surgem com a finalidade de diminuir a sobrecarga atual de dados na *Web* através da seleção de conteúdo baseada nas preferências do usuário. De acordo com Figueira (2008), eles são alvos de pesquisa no campo da recuperação da informação, mas podem abordar pesquisas mais amplas para estimar sua recomendação, seja através de dados semânticos ou da análise de redes sociais. Desse modo, o grande desafio desse modelo é unir a estrutura e o significado semântico dos dados. Uma vez que o ator principal em um sistema de recomendação é o usuário, é fundamental que o sistema não apenas crie elos entre o sujeito e a informação, mas sim que lhe faça entender tais elos.

No contexto das receitas, não foram encontrados pelo grupo, sistemas de recomendação com que se proponham a fornecer ao usuário recomendações de ingredientes que tornem suas receitas mais saudáveis. Existem atualmente portais de comércio eletrônico que utilizam a técnica de recomendação, como a Amazon.com e a iTunes.com. Contudo, nestes casos, a recomendação é estimada a partir de preferências do usuário no site, enquanto o Tudo Saudável irá recomendar ingredientes a partir de seus respectivos níveis calóricos, não levando em conta, num primeiro momento, a qualidade da substituição.

Somente depois de feito esse processo, o usuário poderá melhorar a recomendação observando outros requisitos além do valor calórico, o que será realizado através do *feedback* dado pelo sujeito, logo após ter acessado uma recomendação.

Sendo assim, o desenvolvimento de um sistema de recomendação *Web* baseado em indicação de receitas mais saudáveis, se confirma um processo inovador e contemporâneo. A disponibilização do Tudo Saudável no ambiente da *Internet* garante a conveniência de levar ao usuário final as recomendações de forma cômoda e prática. Tudo isso, acaba incentivando o usuário final a adotar hábitos alimentares que beneficiarão a sua saúde e bem-estar.

2.3. Métricas para Análise de Dados

A fim de estabelecer relações em grandes redes, como sites de compras ou redes sociais, foram definidas certas métricas para analisar dados. Essas métricas ajudam a estudar grandes redes de forma a extrair-se informações úteis, como padrões, semelhanças e preferências. Assim, métricas para análise de dados são ferramentas essenciais para trabalhos que envolvam redes de sistemas de recomendação.

Pode-se classificar uma rede como um conjunto de “vértices” interligados entre si. Esses vértices comumente possuem relações que podem ser representadas por grafos. Tais grafos evidenciam possíveis padrões e semelhanças de dados pertinentes à rede. Esses dados podem ser estudados e utilizados para se obter informações importantes sobre a rede e suas relações, possibilitando um novo mundo de opções.

Um componente em um grafo é um conjunto de vértices, em que cada vértice possui um caminho para todos os outros vértices. Um componente é chamado de fortemente conectado (*SCC - Strongly Connected Component*) quando existe um caminho direcionado entre cada par de vértices do conjunto. Um componente é fracamente conectado (*WCC - Weakly Connected Component*) se o caminho não é direcionado.

A classificação de grafos em *SCC* ou *WCC* é pertinente para o uso de métricas pela distância média e pelo diâmetro de um grafo. A distância média de um grafo é a média das arestas nos caminhos mínimos existentes entre os pares de vértices do grafo. Normalmente, a distância média é computada apenas no *SCC* para grafos direcionados ou no *WCC* para grafos não direcionados pois caminhos entre vértices em componentes diferentes são inexistentes.

Já o diâmetro é definido como a distância do maior caminho mínimo existente no grafo e, em geral, é também computado somente para vértices do *WCC* ou do *SCC*.

Se uma rede possuir diâmetro pequeno e um alto coeficiente de agrupamento ela pode ser considerada uma rede *small-world*. Estas propriedades foram verificadas em várias redes como a *Web* e redes sociais *online*, como o *Orkut* e o *YouTube*. O conceito de redes *small-world* foi definido após o experimento de Milgram e seu resultado que ficou conhecido como o princípio dos seis graus de separação.

Uma rede pode ser então, modelada por um grafo, que, por sua vez, pode ser caracterizado a partir de diversas métricas. Uma dessas métricas consiste na análise do expoente α (de $k^{-\alpha}$), obtido através de uma regressão linear, em que $k^{-\alpha}$ é proporcional à probabilidade de um vértice ter grau k . O *Gnuplot* e o *Matlab* são ferramentas utilizadas para realizar a regressão e calcular o valor de α . A distribuição dos graus dos vértices de uma rede já foi caracterizada em várias redes como redes de *e-mails*, a *Web* e redes neurais, seguindo uma lei de potência. Essas redes cujas distribuições dos graus dos vértices seguem uma lei de potência são chamadas redes *power-law*.

Quando a rede segue leis de potência caracterizadas por vértices de grau alto que tendem a se conectar aos outros vértices de grau alto ela é conhecida por rede livre de escala (*scale free*).

A assortatividade ou disassortatividade de uma rede são ainda conceitos importantes, geralmente estimados avaliando os valores de $knn(k)$ em função de k , em que $knn(k)$ é o grau médio de todos os vizinhos dos vértices com grau k . Segundo Newman (2002), assortatividade é uma medida típica de redes sociais. Valores crescentes indicam assortatividade, isto é, vértices com graus maiores tendem a se conectar a vértices com um número maior de conexões. Valores decrescentes de $knn(k)$ em função de k , por sua vez, indicam uma rede disassortativa.

Uma métrica muito utilizada é o coeficiente de agrupamento de uma rede (*clustering coefficient*). O CC de uma rede representa a densidade das arestas entre os vizinhos de um vértice e é calculado como a média dos coeficientes de agrupamento de todos os seus vértices. O cálculo ($CC[I]$) de um vértice é feito descobrindo-se a razão entre o número de arestas existentes entre os vizinhos de um vértice I e o número máximo de arestas possíveis entre estes vizinhos, como mostra a Figura 2. Desta forma, se temos que todas as três arestas possíveis entre os vizinhos de um vértice I de fato existem, $CC[I] = 1$; se temos que exista apenas uma destas arestas, $CC[I] = 1/3$, e assim por diante.

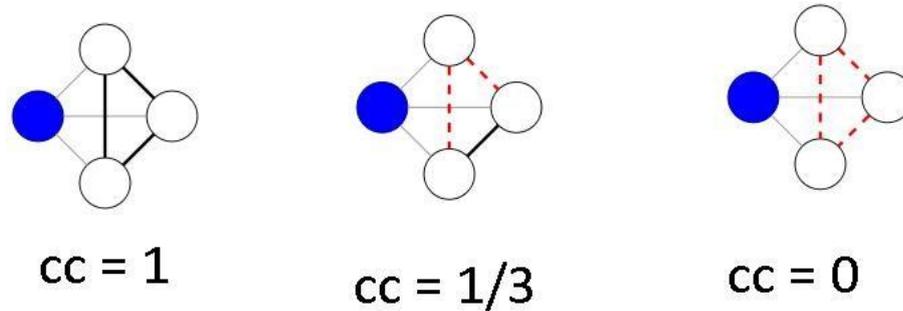


FIGURA 2: Cálculo do Coeficiente de Agrupamento de um Nodo em Três Cenários Diferentes

Uma forma interessante de se observar a reciprocidade de um vértice I em um grafo direcionado é medindo a porcentagem dos vértices apontados por I que apontam para ele.

Outra métrica interessante é o coeficiente de reciprocidade ρ , uma métrica que captura a reciprocidade das interações da rede. O coeficiente de reciprocidade ρ é definido pelo coeficiente de correlação entre entidades da matriz de adjacência representativa do grafo direcionado. Em outras palavras, seja a matriz a definida como $a_{ij} = 1$ se há uma aresta de i para j no grafo, e seja definida como $a_{ij} = 0$, caso não haja. A reciprocidade da rede ρ é:

$$\rho = \frac{\sum_{i \neq j} (a_{ij} - \bar{a})(a_{ji} - \bar{a})}{\sum_{i \neq j} (a_{ij} - \bar{a})^2},$$

onde o valor médio $\bar{a} = \sum_{i \neq j} a_{ij} / N(N-1)$ e N é o número de usuários no grafo. O coeficiente de reciprocidade indica se o número de arestas recíprocas na rede é maior ou menor do que o de uma rede aleatória. Se o valor ρ é maior do que 0, a rede é recíproca; caso contrário, ela é anti-recíproca.

Já o *PageRank* é um caso interessante. O seu algoritmo assinala um peso numérico para cada vértice com o propósito de estimar sua importância relativa no grafo. O algoritmo foi proposto pela primeira vez por *Brin and Page* para ordenar resultados de busca do protótipo de máquina de busca da *Google*. Sua lógica é que uma página *Web* é importante se existem muitas páginas ou páginas importantes apontando para ela. A equação que calcula o *PageRank* (PR) de um vértice I ($PR[I]$), é definida como:

$$PR(i) = (1-d) + d \sum_{v \in S(i)} \frac{PR(v)}{N_v}$$

2.4. Coletor de Dados

Mineração de dados nada mais é que o processo de coleta de informações acionáveis em grandes bases de dados. Segundo Liddy (2000), tal processo consiste em três etapas:

- Coleta dos Dados: na qual ocorre a seleção e pré-processamento dos dados que formarão a base textual de trabalho a ser processada. É nesse momento que ocorre a escolha de sites ou outras fontes de texto e informações.
- Processamento dos Dados: fase responsável pela utilização do algoritmo de mineração de dados para processar os dados já preparados na fase anterior.
- Análise dos Dados: etapa onde os dados processados são submetidos a técnicas que permitirão a utilização direta da informação ou, se for o caso, a análise humana por métodos manuais.

Com isso, a primeira fase de desenvolvimento do Sistema de Recomendação Receitas Saudáveis – Tudo Saudável, fez uso desse processo para a construção da base de ingredientes, que se firmou como diretriz para a formulação das recomendações. Os dados foram coletados do site Tudo Gostoso, processados por scripts de mineração e, por fim, analisados e desambiguados manualmente.

2.5. Bancos de Dados Não Relacionais

Desde sua criação no início dos anos 1970, o Modelo Relacional de dados tem sido utilizado em larga escala pela grande maioria dos sistemas de gerenciamento de banco de dados (YANK, 2004). Tendo surgido como sucessor dos modelos hierárquico e de rede, o modelo relacional tornou-se padrão para a grande maioria dos SGBDs (Sistemas Gerenciadores de Banco de Dados). Seus elementos básicos são as relações, ou tabelas, as quais são compostas de linhas, ou tuplas, e colunas, ou atributos.

Uma característica importante desse modelo é a utilização de restrições de integridade. As restrições de integridade mais comuns são as chaves, mais especificamente, as chaves primárias e as chaves estrangeiras.

A chave primária tem o objetivo de assegurar a identificação única das tuplas das tabelas. A chave estrangeira torna os valores de determinado atributo dependentes dos valores existentes em outro atributo, normalmente de outra tabela. Outra característica importante do Modelo Relacional é o processo de Normalização. Seu objetivo é a aplicação de certas regras sobre as tabelas do banco de dados, de forma a garantir o projeto adequado dessas tabelas.

A utilização de tais recursos facilitou a vida dos desenvolvedores de aplicações, possibilitando que estes pudessem se preocupar exclusivamente com o foco da aplicação. Todos esses recursos, porém, ajudaram a manter os SGBDs Relacionais em posição de destaque entre os mais diversos tipos de ambientes computacionais, mas não impediram o

surgimento de certos problemas, principalmente devido ao crescimento vertiginoso do volume de dados presentes nos bancos de certas organizações.

Devido ao forte crescimento do número de aplicações, soluções, recursos e tudo o que se refere a sistemas computacionais nas últimas décadas, o volume de dados associados a tais tipos de sistemas também teve um ritmo de crescimento acelerado.

De um modo geral, os principais problemas encontrados com a utilização do Modelo Relacional estavam concentrados na dificuldade em se conciliar tal modelo com a demanda por escalabilidade cada vez mais frequente.

Segundo Leavitt (2010), devido a sua natureza estruturada, os desenvolvedores começaram a perceber a dificuldade em se organizar os dados do Modelo Relacional em um sistema distribuído trabalhando com particionamento de dados. É justamente nesse ponto em que o foco das soluções não-relacionais estão concentradas.

Assim como o termo não-relacional, o termo NOSQL não ajuda a definir o que esses bancos são de fato. O termo NOSQL tem sido usado com o significado de “Não apenas SQL” numa tentativa da comunidade de reconhecer a utilidade dos modelos tradicionais e não divergir as discussões, sendo utilizado principalmente em casos em que o Modelo Relacional não apresentava performance adequada.

O propósito, portanto, das soluções NOSQL não é substituir o Modelo Relacional como um todo, mas apenas em casos nos quais seja necessária uma maior flexibilidade da estruturação do banco. Em português já existe o acrônimo MRNN para Modelo Relacional Não-Normalizado.

Quando se analisa a possibilidade de se optar por uma estratégia NOSQL em detrimento de um SGBD tradicional, é preciso levar em consideração algumas questões básicas, como, por exemplo, os critérios de escalonamento, consistência de dados e disponibilidade.

A questão do escalonamento, por exemplo, é essencial porque é justamente nesse ponto em que os bancos NOSQL apresentam as principais vantagens em relação aos SGBDs relacionais principalmente por terem sido criados para esse fim, enquanto os sistemas relacionais possuem uma estruturação menos flexível e menos adaptada para cenários em que o escalonamento se faz necessário.

Os principais benefícios do uso dos NOSQL, portanto, podem ser identificados como: maior disponibilidade, menor tempo de resposta para consultas, paralelismo de atualização de dados e maior grau de concorrência.

Os bancos de dados NOSQL foram especialmente projetados para atender essas características de forma mais natural. Ao se substituir um SGBD relacional por uma solução

NOSQL, a arquitetura perde em consistência, mas pode ganhar em flexibilidade, disponibilidade e performance.

3. Projeto do Sistema

A seguir serão detalhadas as funcionalidades do Sistema de Recomendação de Receitas Saudáveis – Tudo Saudável. Consistem basicamente em, cadastros e buscas nas tabelas principais. Serão apresentadas todas as funções do sistema, entretanto essas são limitadas de acordo com o tipo de acesso do Usuário. Além disso, abaixo é exposto as etapas de coleta e tratamentos dos dados.

3.1. Definição das Funcionalidades

O Sistema de Recomendação de Receitas Saudáveis consiste em uma forma de proporcionar aos usuários de sites culinários uma maneira de buscar receitas culinárias menos calóricas. Para tal, o sistema cumpre em sua versão *Web*, um conjunto de funcionalidades básicas para um ator principal: o Usuário, que é o público alvo de sites de culinária e utilizador mais leigo.

Ao Usuário são oferecidas pelo sistema as seguintes funcionalidades:

- Realizar o cadastro/login na página destinada, onde o Usuário pode criar e ter acesso a uma conta única para usufruir do sistema em sua integridade – é exigido um login para cadastrar uma receita, avaliar e comentar as substituições;
- Inserir uma receita no Sistema, bem como fazer alterações e exclusões das mesmas;
- Apresentar ao Usuário a receita pesquisada a partir do nome do prato ou, ainda, as receitas que tenham um ou mais ingredientes buscados pelo Usuário;
- Informar ao Usuário opções de substituições de ingredientes saudáveis, apresentando um pequeno quadro com os ingredientes em questão e seus respectivos níveis calóricos;
- Receber avaliações e críticas dos Usuários sobre as recomendações de substituições em indicadores de satisfação e campos de comentários.

3.2. Coleta, análise e desambiguação dos ingredientes das Receitas

A coleta de receitas para o projeto foi feita sobre a categoria Bolos e Tortas do site Tudo Gostoso. Este processo de coleta se deu através de script que será apresentado no Apêndice A. A opção por essa categoria em especial, foi feita baseando-se principalmente na vasta quantidade de receitas encontradas nela. Além disso, o fato desta categoria possuir um grande número de ingredientes distintos, cerca de 603 ao todo (FERREIRA. Willyan M., 2013), também influenciou na escolha.

A Tabela 1 abaixo indica, de forma mais concisa, a viabilidade de se trabalhar com esta categoria pelo seu grande número de receitas:

Nome da Categoria	Número de Receitas
Bolos e Tortas	30.693
Carnes	7.987
Aves	6.936
Peixes e Frutos do Mar	6.005
Saladas, Molhos e Acompanhamentos	21.595
Sopas	3.280
Massas	19.089
Bebidas	5.492
Doces e Sobremesas	36.526
Lanches	20.368
Prato Único	8.755

TABELA 1: Receitas por Categoria no Site Tudo Gostoso

O coletor de dados, como já foi explicitado no referencial teórico, é responsável por buscar as informações a serem trabalhadas. Como o site Tudo Gostoso não possui uma *Application Programming Interface* (API), a coleta se deu de forma automática a partir de todo o código fonte das páginas onde estão expostas as receitas através de scripts.

O mecanismo foi desenvolvido em Python, linguagem que facilita o trabalho com grande volume de dados semiestruturados ou não estruturados. A partir do código HTML, foram coletadas as sessões dos ingredientes, os quais foram posteriormente identificados.

Após a etapa de coleta, ocorreu a identificação dos ingredientes que consistiu, inicialmente, em retirar a parte de medidas, unidades e stopwords (o conectivo das sentenças) de cada ingrediente. Essa identificação se deu pelo script que é exposto no apêndice B. Assim, no exemplo da Figura 3, “1 xícara de leite”, o trabalho consistia em retirar apenas o ingrediente “leite”:

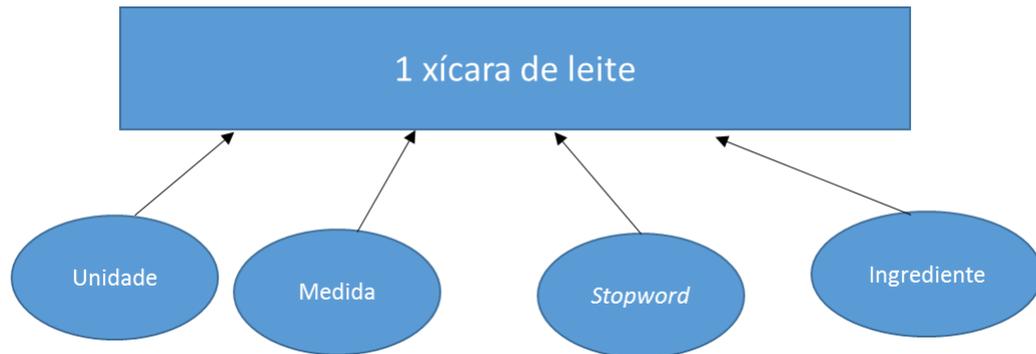


FIGURA 3 : Identificação dos ingredientes

Depois da identificação, os ingredientes passaram por uma fase de desambiguação para que pudessem servir como base de dados. Toda a etapa de desambiguação foi feita através de script¹. Por ser um sistema colaborativo de receitas, no qual os usuários têm liberdade para alimentar a base de dados, que não é filtrada, o Tudo Gostoso pode apresentar desafios para a coleta. Falhas como erros ortográficos, erros de digitação, regionalismo de ingredientes, ingredientes com nomes estrangeiros, além de agrupamento de ingredientes, acabam dificultando o processo. No exemplo abaixo, erros de ortografia foram identificados e corrigidos:

Oio

Olieo

Olheo

→

Óleo

Para validar a identificação de ingredientes foram escolhidas aleatoriamente 100 receitas da base, essas foram submetidas ao script de identificação e passaram também por uma identificação manual. O resultado desse experimento, foi uma taxa de acerto 97% das receitas e apenas 6 ingredientes identificados de forma incorreta.

¹https://www.dropbox.com/s/nozcsqktq3f7tsm/correcao_base.py

3.3. Pesquisa de níveis calóricos dos ingredientes

Foi feita uma pesquisa baseada nos ingredientes coletados na primeira funcionalidade afim de obter seus níveis calóricos para formular as recomendações. A pesquisa foi obtida usando-se fontes confiáveis na *Internet*, especialmente tabelas nutricionais. De forma semelhante a coleta dos ingredientes realizada inicialmente, foi traçado um paralelo entre uma fonte da *Internet* e a lista dos ingredientes já coletados e corrigidos.

A principal fonte de informações para esta etapa foi o site Boa Forma, selecionado pela grande quantidade de ingredientes com seus respectivos níveis calóricos pesquisados encontrados, além da organização simples e direta de suas tabelas nutricionais, o que facilitou a coleta.

3.4. Formulação das recomendações

Após a análise dos ingredientes e a pesquisa nutricional, foram formuladas pela equipe as recomendações de substituições de ingredientes. Os critérios escolhidos para as recomendações foram o nível calórico do ingrediente proposto e sua viabilidade na receita, já que a mesma não poderá perder suas propriedades originais como textura, aspecto e consistência. Neste procedimento foram levados em conta tanto os grafos obtidos com as relações das redes de ingredientes extraídas das receitas do site Tudo Gostoso no artigo Comer, Comentar e Compartilhar, quanto as próprias opiniões dos usuários registradas nos comentários. Assim, tendo sempre em mente a redução de calorias que a substituição deve acarretar, foram traçadas recomendações como as do exemplo:

Açúcar Refinado (387 kcal – 100 g)	→	Adoçante (40 kcal – 100 g)
Leite Condensado (744 kcal – 240 g)	→	Leite Integral (146 kcal – 240 g)
Pão francês (135 kcal – 50 g)	→	Pão de Forma (124 kcal – 50 g)

3.5. Estruturação do banco de dados

Nessa etapa, todos os dados coletados e analisados, assim como suas atribuições que foram encontradas, foram armazenados em um banco de dados. Apesar dos bancos de dados não relacionais serem uma alternativa aplicável à recuperação da informação, o sistema de gerenciamento de banco de dados (SGBD) escolhido foi o *MySQL* através da ferramenta *PhpMyAdmin*, devido principalmente à sua facilidade de manuseio.

O *MySQL* utiliza a linguagem SQL (Linguagem de Consulta Estruturada, do inglês *Structured Query Language*) como interface. De acordo com Suehring (2002), é atualmente um dos bancos de dados mais populares, com mais de 10 milhões de instalações pelo mundo. Como foi escolhido justamente devido a prática que a equipe já possuía, a estruturação do banco de dados ocorreu sem grandes dificuldades.

Já o *PhpMyAdmin*, é uma ferramenta de manipulação do *MySQL*. É muito útil, principalmente, para desenvolvedores *Web*.

3.6. Desenvolvimento da Interface *Web*

Completadas as etapas anteriores, foram desenvolvidas as telas do software para a *Web*. O site do Sistema de Recomendação de Receitas Saudáveis – Tudo Saudável, apresenta basicamente um campo de pesquisa onde podem ser feitas as buscas por receitas ou ingredientes e uma página de login e entrada para os usuários. No site, pode-se também cadastrar uma receita e efetuar uma avaliação e comentários acerca das substituições, porém os usuários devem estar logados em suas respectivas contas para executarem tais ações. Além disso, o site conta com um espaço para apresentar informações e curiosidades sobre o projeto.

Todas as telas foram desenvolvidas visando especificamente o ambiente *Web* utilizando o *framework Bootstrap*. Esse é uma coleção de ferramentas, livre e de código aberto, para a criação de sites e aplicações *Web*. Além dos elementos HTML regulares, o *Bootstrap* contém outros elementos da interface vulgarmente utilizados. Estes incluem botões com funções avançadas (por exemplo, agrupamento de botões ou teclas com a opção *drop-down* e listas de navegação, guias horizontais e verticais, navegação, navegação de trilha, paginação, etc.), rótulos, recursos tipográficos avançados, miniaturas, mensagens de aviso e uma barra de progresso.

Os componentes são implementados como classes CSS, que devem ser aplicadas a certos elementos HTML em uma página. Também são disponibilizadas extensões JavaScript opcionais. É conhecido principalmente a facilitar o desenvolvimento de sites dinâmicos e aplicações *Web*.

4. Projeto Conceitual

A seguir estão o diagrama de Caso de Uso, a documentação do ator e o diagrama de Implantação com a fim de ilustrar a arquitetura empregada e facilitar o entendimento.

4.1. Diagrama de Caso de Uso

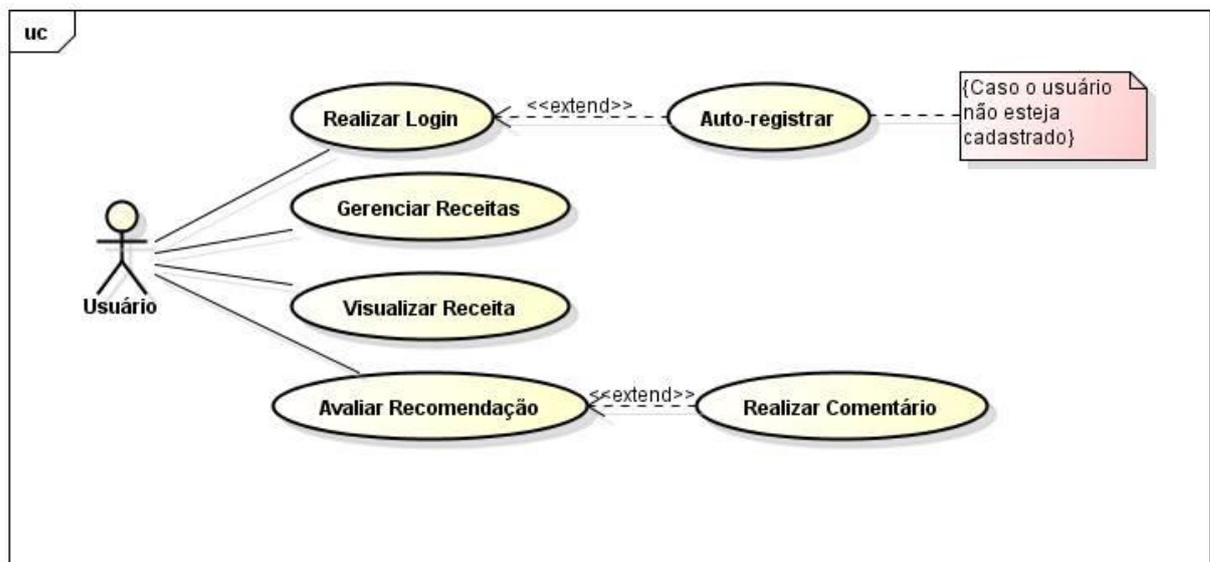


FIGURA 4: Diagrama de Caso de Uso

O diagrama de Caso de Uso ilustrado na Figura 4 acima, representa as ações do ator envolvido Sistema. O Usuário possui a opção de login, e caso não esteja cadastrado ele poderá efetuar seu auto registro. Além disso, o ator pode visualizar as receitas ou gerenciá-las e avaliar as substituições e realizar comentários. No Apêndice c, será apresentado a descrição de tal Caso de Uso.

4.2. Documentação do Ator

No tópico seguinte, será melhor detalhado o ator que interage com o Sistema.

4.1.1.1. Usuário

Por ser um Sistema voltado para plataforma *Web*, há uma grande amplitude de usuários, mas espera-se que seja procurado, principalmente, por pessoas interessadas em praticar hábitos alimentares mais saudáveis, atrelados ou não a perda de peso e melhora da imagem estética.

O Usuário possui a ação de logar em sua conta de acesso do sistema ou, caso não cadastrado, se auto registrar para obter integralmente as vantagens do sistema. Tal ator possui também a alternativa de cadastrar receitas, bem como alterá-las e excluí-las. Além disso, o Usuário pode visualizar as receitas já disponibilizadas e que possuem substituições já formuladas. Por fim, ele pode, se desejar, avaliar a substituição feita e, se achar pertinente, realizar um comentário elogiando, criticando ou sugerindo modificações.

4.3. Diagrama de Implantação

A Figura 5 apresenta o diagrama de implantação do Sistema, cujo intuito é representar o ambiente físico no qual ele será implantado. Por ser um sistema *Web*, o ambiente de execução do sistema é a *Internet* e os dispositivos necessários para seu funcionamento são o computador pessoal, o servidor *Web* intranet e o servidor de banco de dados.

Os ambientes de execução no computador pessoal são as plataformas *Windows* e *Linux*, e os artefatos para execução podem ser *Google Chrome*, *Mozilla Firefox* ou *Internet Explorer*. Já no servidor *Web* intranet, o ambiente de execução é o *Windows*, mais especificamente através do do servidor *Web Apache*. Por fim, encontra-se no servidor de banco de dados o ambiente *Windows* mais uma vez, por intermédio do banco de dados *MySQL*.

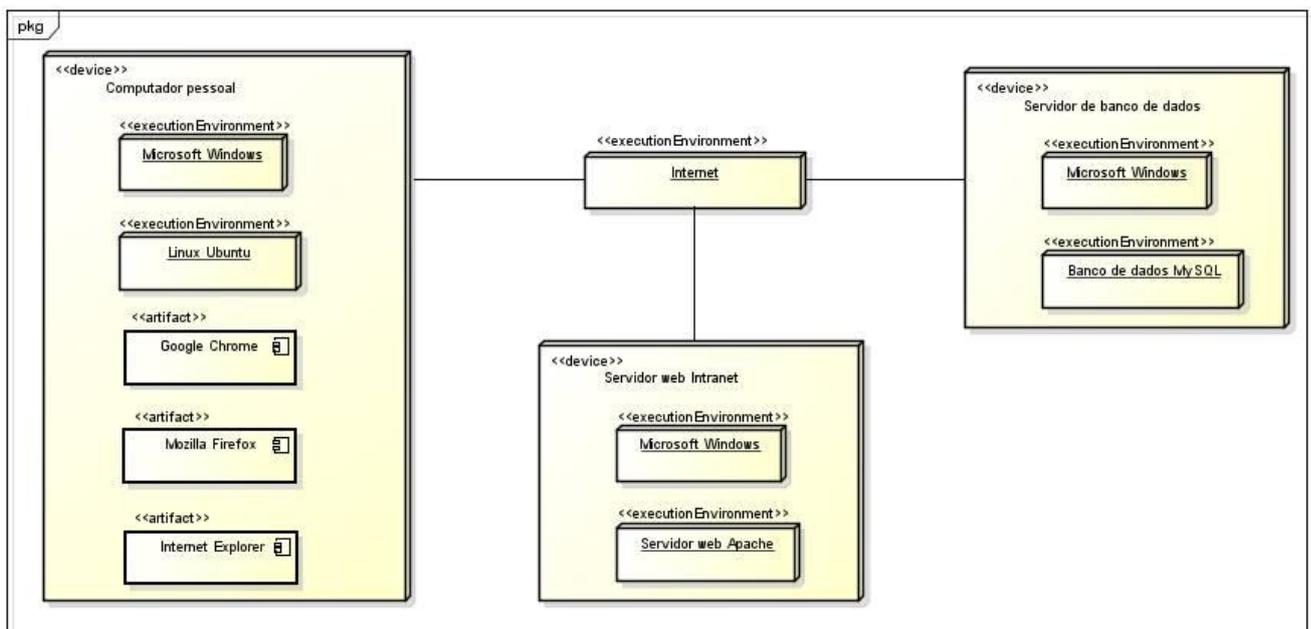


FIGURA 5: Diagrama de Implantação

4.4. Diagrama de Classe

Na Figura 6, é apresentado o Diagrama de Classe e seus relacionamentos do Sistema de Recomendação de Receitas Saudáveis – Tudo Saudável.

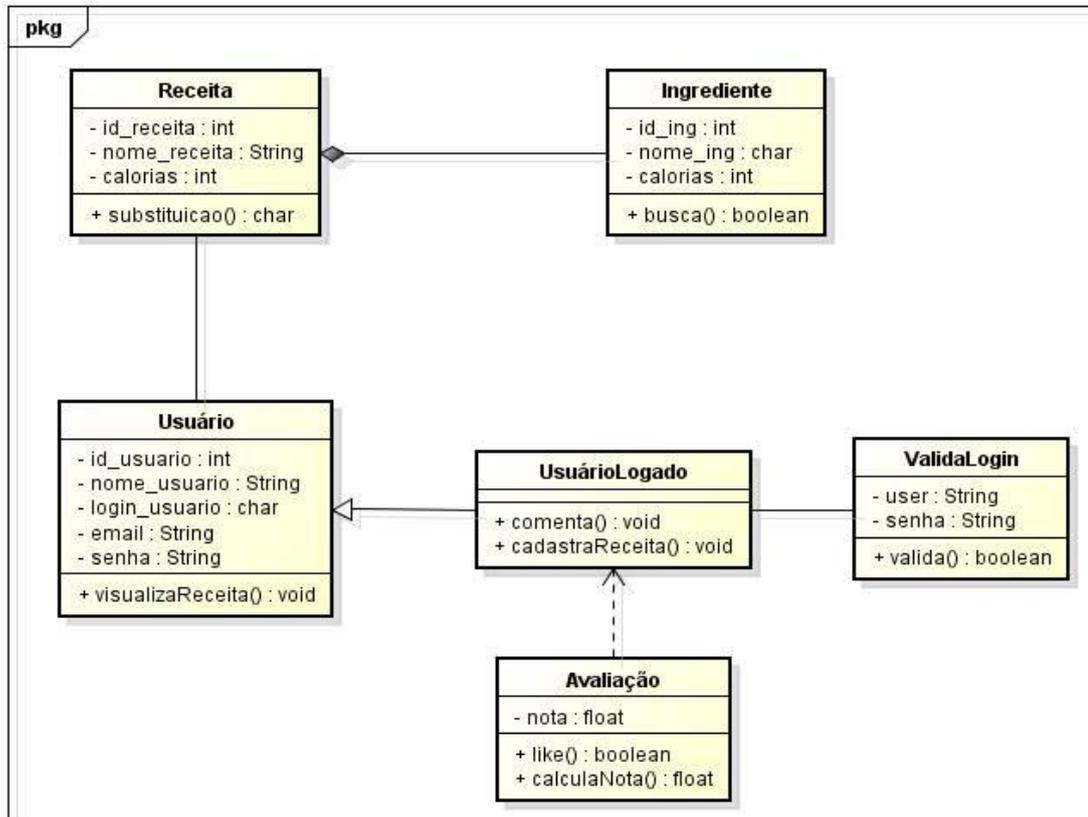


FIGURA 6: Diagrama de Classe

5. Projeto Físico

Abaixo é exposto o DER – Diagrama de Entidade e Relacionamento e o DTR – Diagrama de Tabelas Relacionais que ilustram as entidades e relacionamentos que compõem o banco de dados do Sistema de Recomendação de Receitas Saudáveis - Tudo Saudável.

5.1. DER – Diagrama de Entidade e Relacionamento

O modelo ilustrado na Figura 7 representa as entidades relacionais do Sistema. Ele é composto pela entidade Usuário que possui sua identificação, nome, login, e-mail e senha. Além disso, a entidade Usuário se comunica com as de Receita e Ingredientes, já que esse pode cadastrar uma receita ou pesquisá-las por nome da receita ou pelos ingredientes que a

compõem. Outra funcionalidade que o usuário possui é de avaliação caracterizada pela entidade Comentário na qual ele pode avaliar com uma nota ou comentar.

Os ingredientes que possuem uma recomendação, se relacionam com a entidade Substituição. Como um mesmo ingrediente pode estar presente em diversas receitas, fez-se necessária a construção da entidade Ing_rec, que impede a restrição de um ingrediente à apenas uma receita.

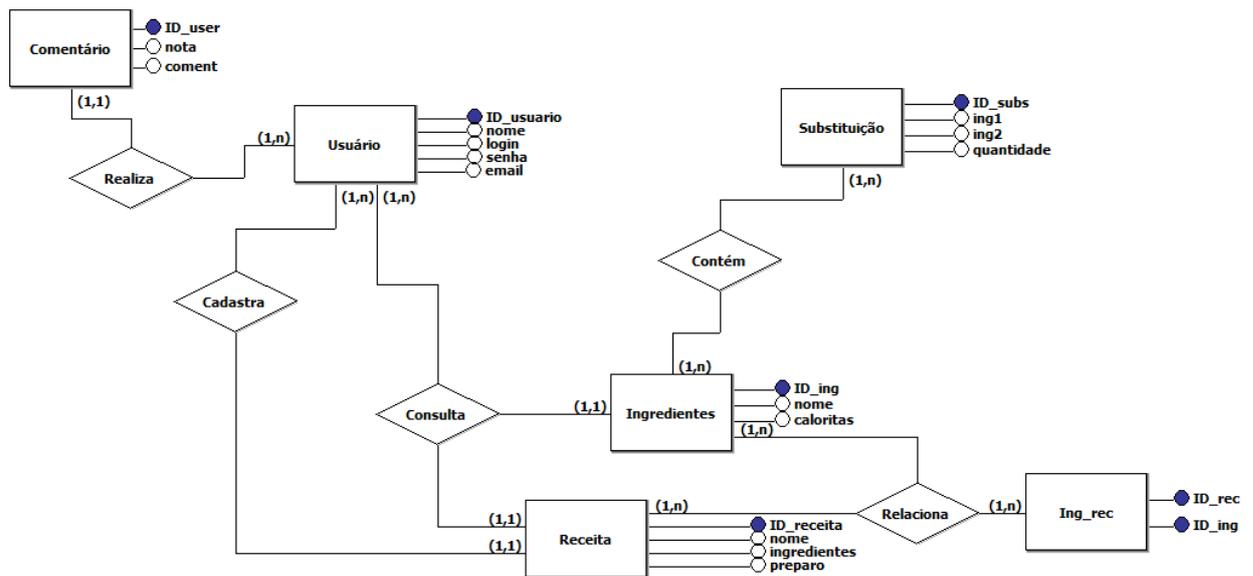
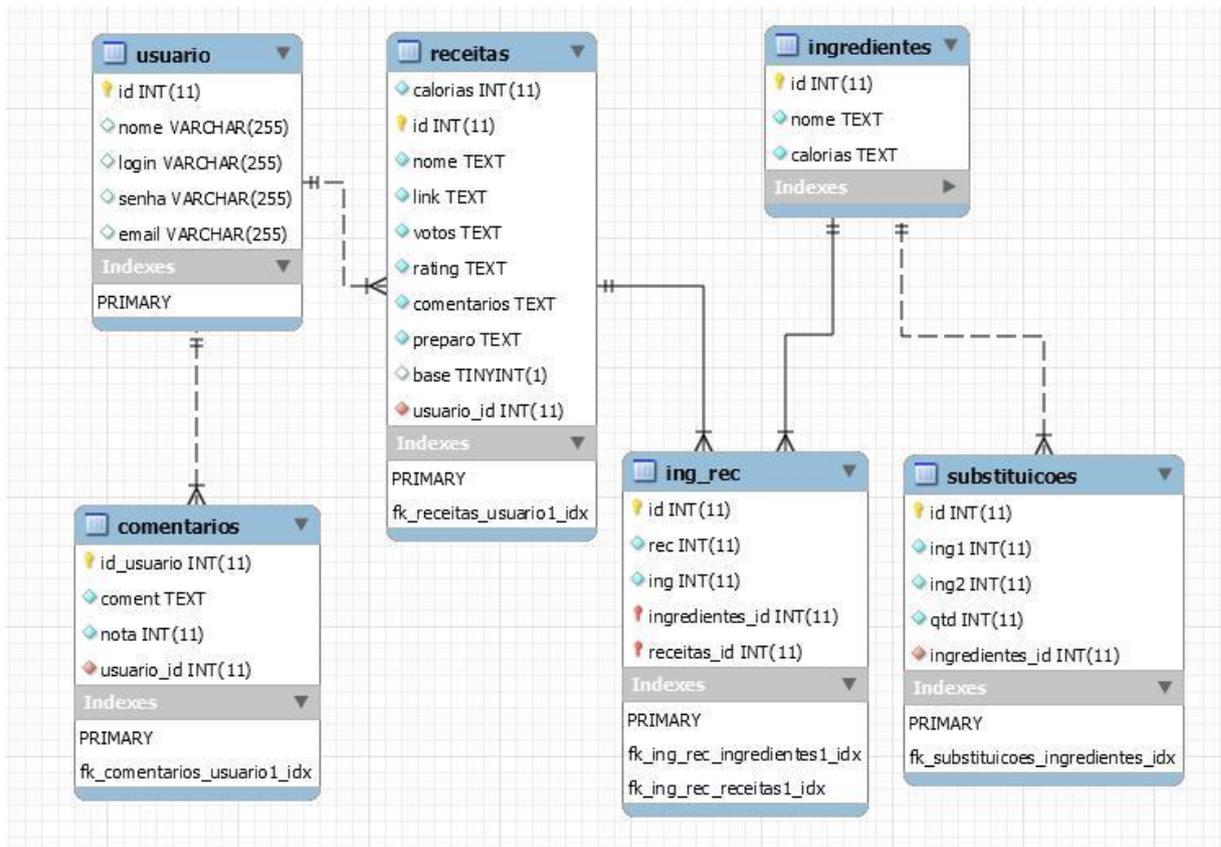


FIGURA 7: Modelo de Entidade e Relacionamento

5.2. DTR - Diagrama de Tabelas Relacionais

A Figura 8 contém o diagrama de tabelas relacionais que demonstra as relações entre



as entidades que compõem o Sistema.

FIGURA 8: Modelo de Entidade e Relacionamento

6. Resultados Finalizados

Na fase inicial do projeto, foram coletados, através de scripts em *Python*, os ingredientes das receitas da categoria Bolos e Tortas do site Tudo Gostoso. Esses foram analisados e desambiguados manualmente para que pudessem servir de base para o banco de dados. Uma vez que os ingredientes foram coletados, foram pesquisados os níveis calóricos. Aqueles que tiveram retorno foram armazenados no banco criado pela equipe. A partir daí, foram formuladas substituições que implicassem em receitas menos calóricas.

Posteriormente, foi desenvolvida a interface *Web* do sistema usando o *framework Bootstrap*. Obteve-se assim, um site limpo, prático e de fácil navegação. A seguir, serão expostas as principais telas do Site Tudo Saudável.

Na Figura 9, está presente a tela inicial do site, onde são encontrados a barra de menu principal e campo para login do usuário

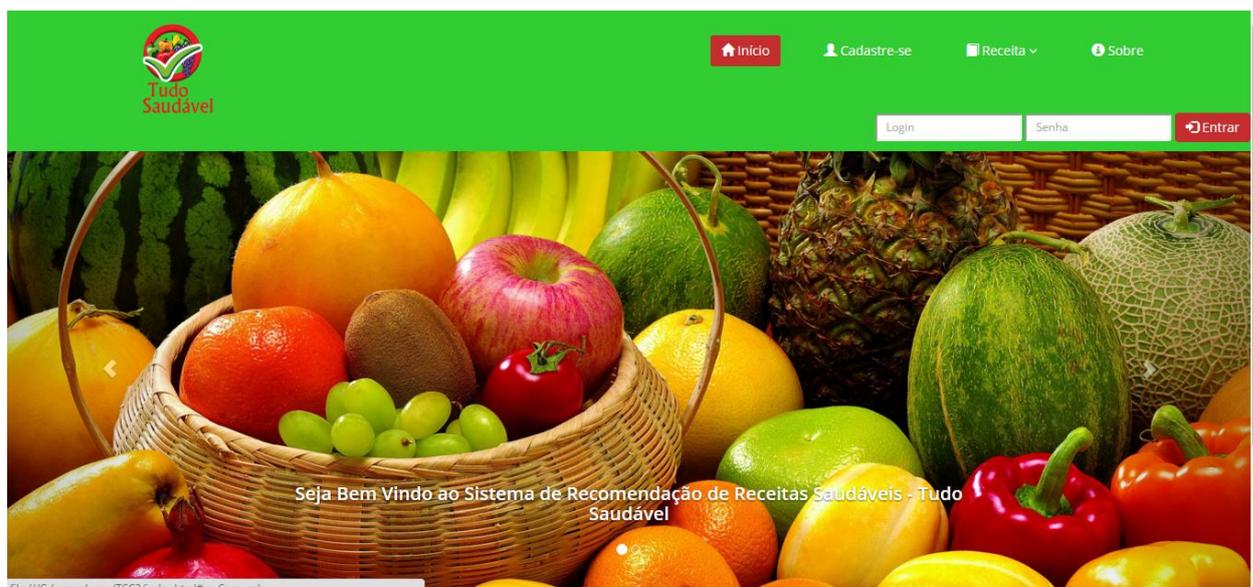


FIGURA 9: Tela Inicial

Caso o usuário ainda não seja cadastrado, tem-se a opção de se auto registrar para melhor uso do site, já que existe algumas restrições quando o usuário não está logado. Tal aplicação pode ser visualizada na Figura 10, onde o usuário insere seus dados para cadastro.

FIGURA 10: Tela de Cadastro do Usuário

Depois de cadastrado, o Usuário também pode modificar dados pessoais como nome, e-mail e senha, como pode ser observado na Figura 11.

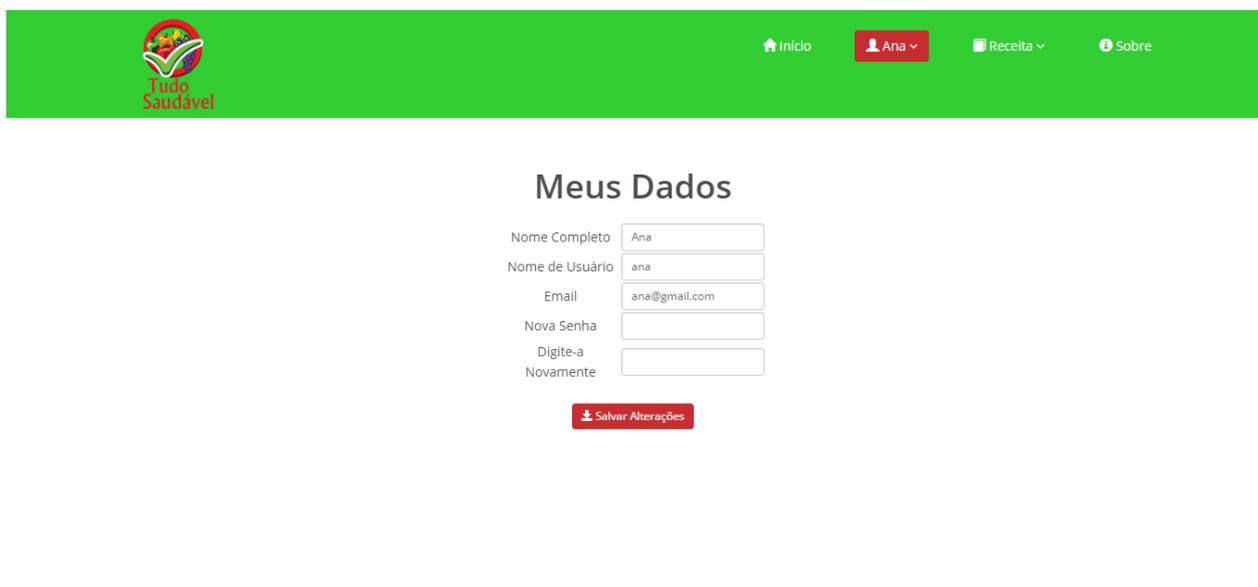


FIGURA 11: Tela de Dados Pessoais do Usuário

Uma das aplicações exclusivas que o usuário logado pode usufruir é o de cadastro de receitas, bem como o de sua manipulação. Abaixo, na Figura 12, tem-se a tela de cadastro de receita, onde o usuário insere o nome da receita, seleciona os ingredientes desejados entre os que já se encontram na base, e suas respectivas medidas. Caso o usuário queira cadastrar um ingrediente diferente que ainda não esteja na base, é lhe dada essa opção. Além disso, é oferecido campo para modo de preparo da receita.



FIGURA 12: Tela de Cadastro de Receita

Outra funcionalidade, é a visualização das receitas que já estão cadastradas no Sistema e, para isso, o usuário não precisa estar logado. Ele pode consultar as receitas por duas formas distintas. A primeira, consiste em buscá-la pelo nome, como pode ser visto na Figura 13.

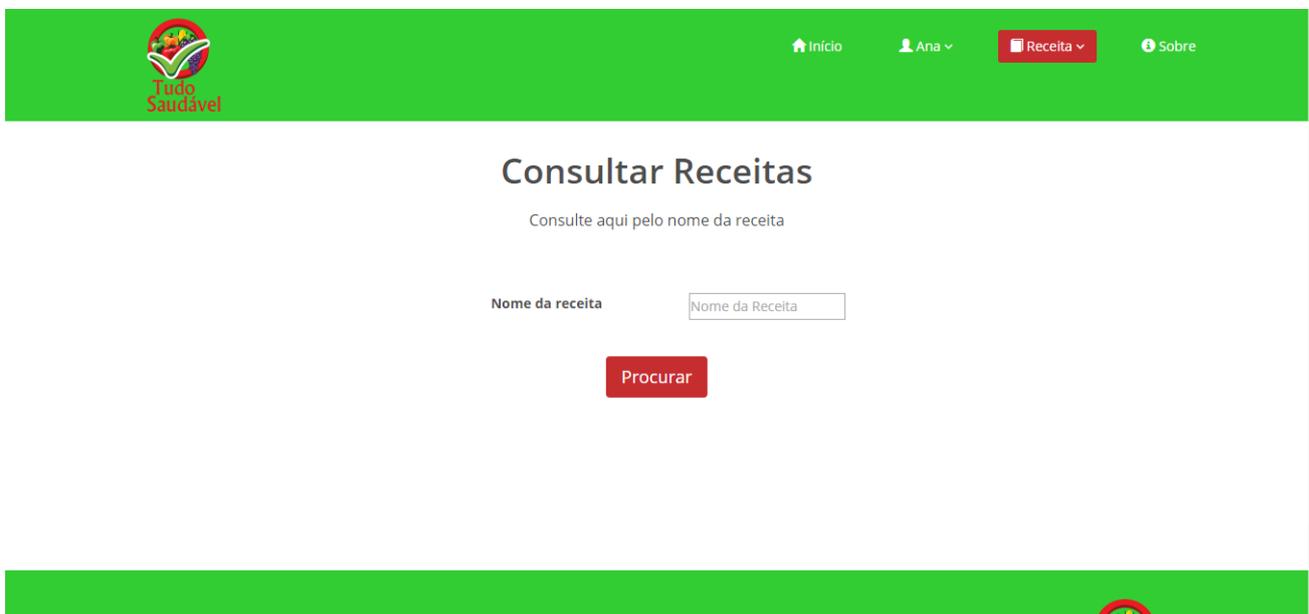


FIGURA 13: Tela de Consulta de Receitas

Já na Figura 14, pode ser vista a segunda forma de buscar receitas no sistema,



selecionando os ingredientes que a compõe.

FIGURA 14: Tela de Consulta de Receitas

Após a solicitação de uma consulta, o sistema devolve uma lista com as receitas que se encaixam nas especificações ditadas pelo usuário, sejam elas pela consulta por nome ou por ingredientes. Isso é visto na Figura 15.

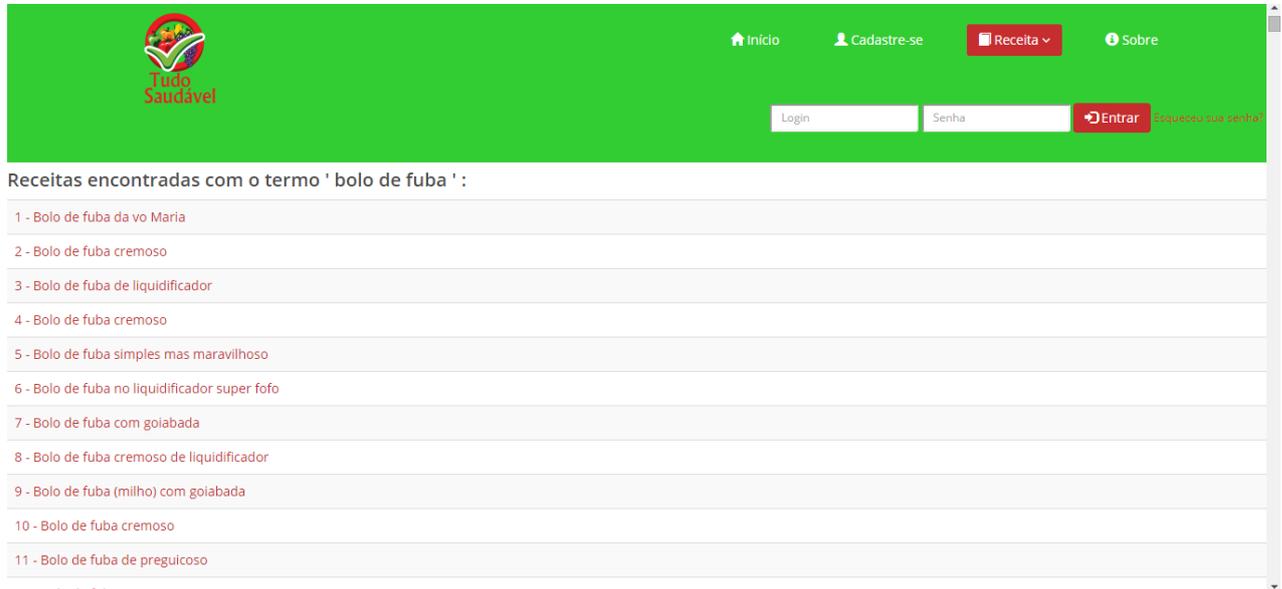


FIGURA 14: Tela de Resultados Encontrados

A partir dos resultados encontrados o usuário pode selecionar a receita buscada por ele e visualizá-la já com a substituição de ingrediente inclusa. A visualização da receita é diferente para usuários não logados e logados. Para os primeiros, a visualização é feita como mostra a Figura 15.

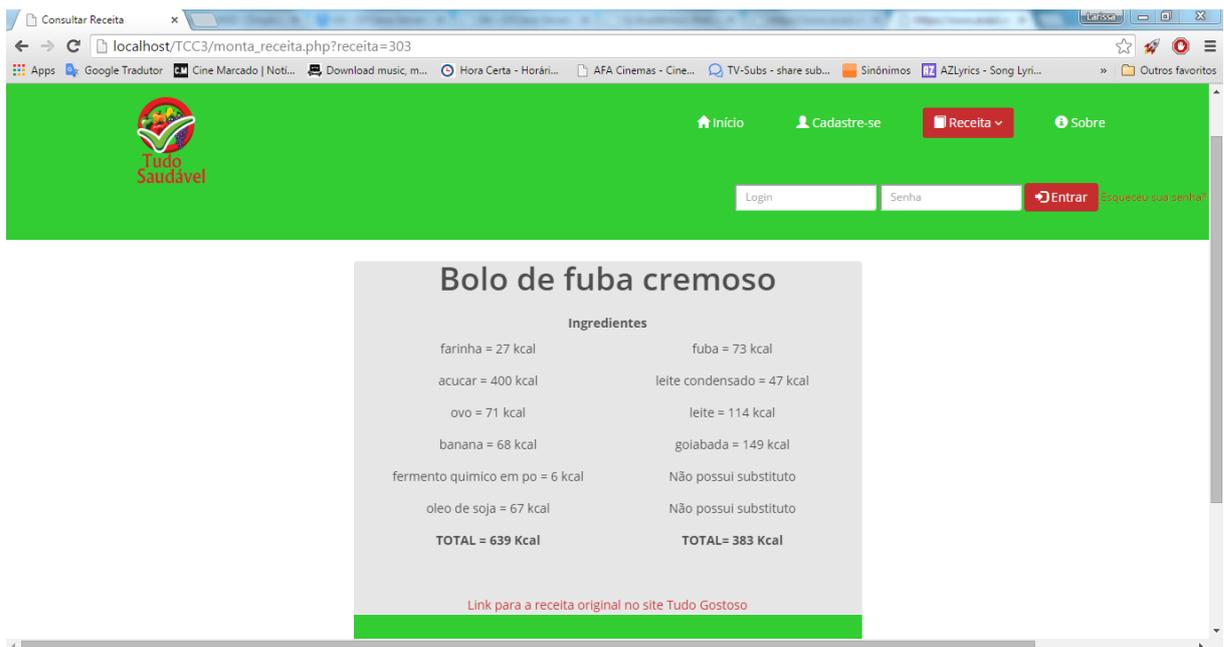


FIGURA 14: Tela de Receitas para Usuários Não Logados

No segundo caso, ou seja, quando é realizado o login, é disponibilizado o recurso de avaliar as substituições através dos botões “like” e “dislike”, bem como o de realizar um comentário expressando sua opinião. Tal funcionalidade pode ser vista na Figura 15, que

Bolo de fuba cremoso

Ingredientes

farinha = 27 kcal	fuba = 73 kcal		
acucar = 400 kcal	leite condensado = 47 kcal		
ovo = 71 kcal	leite = 114 kcal		
banana = 68 kcal	goiabada = 149 kcal		
fermento quimico em po = 6 kcal	NÃO possui substituto		
oleo de soja = 67 kcal	NÃO possui substituto		
TOTAL = 639 Kcal	TOTAL= 383 Kcal		

[Link para a receita original no site Tudo Gostoso](#)

Envie um comentario

Envie um comentario

contém um exemplo de receita.

FIGURA 15: Tela de Receitas para Usuários Logados

Por fim, também é disponibilizado no site do Tudo Saudável uma seção “About”, onde podem ser encontradas informações básicas sobre o projeto, seus objetivos e fundamentos, bem como da instituição pelo qual foi realizado. O acesso a essa página é livre, ou seja, não há necessidade de Login. Na Figura 16 é exposto esse espaço.

TUDO SAUDÁVEL
Sistema de recomendação de receitas saudáveis

Sistema de Recomendação de Receitas Saudáveis - Tudo Saudável é um projeto de conclusão de curso desenvolvido por alunas do 3º ano do curso técnico integrado em informática do CEFET-MG Campus V. O trabalho tem por proposta apresentar substituições de ingredientes viáveis a partir dos ingredientes usados nas receitas de base extraídas do site Tudo Gostoso. Essa base foi estruturada através de um coletor de dados que, segundo Liddy (2000), é um processo de análise natural de ocorrência de texto com a finalidade de descobrir e capturar informações semânticas. Com as correlações dos ingredientes e seus respectivos níveis calóricos, são indicadas ao usuário as substituições mais adequadas.

CEFET-MG
Centro Federal de Educação Tecnológica de Minas Gerais

O CEFET-MG é uma instituição Federal de Ensino Superior (IFES), que atua na Educação Tecnológica de forma verticalizada, em todos os níveis e graus de ensino, da pesquisa aplicada à extensão. Sua função social é promover a formação do cidadão - profissional qualificado e empreendedor - capaz de contribuir ativamente para as transformações do meio ambiente e da sociedade, aliviando a vivência na educação tecnológica e o crescimento do ser humano, consciente e criativo, visando o desenvolvimento econômico e social do País. Além da sede em Belo Horizonte, o CEFET-MG possui outros Campi.

FIGURA 16: Tela de Informações sobre o Projeto

7. Considerações Finais

O presente projeto teve como diretriz fundamental a recuperação de dados alinhado com sistemas de recomendação *online*. Sendo assim, a equipe realizou um extenso trabalho sobre o funcionamento de mecanismos de mineração e análise de dados empregados no desenvolvimento do Sistema.

Acrescentando a isso pesquisas acerca da crescente preocupação dos brasileiros com a saúde alimentar, viu-se que o desenvolvimento do Sistema de Recomendação de Receitas Saudáveis – Tudo Saudável foi relevante. Os ganhos que tal Sistema pode vir a trazer para a população serão benéficos e refletirão diretamente em melhores hábitos alimentares.

Surgiram, inclusive, durante o desenvolvimento, novas possibilidades de expandir o trabalho realizado aqui para projetos futuros. Isso implicaria em possíveis novas pesquisas acerca do assunto e de novos acréscimos para o Sistema, como um módulo *Mobile*.

Assim sendo, percebeu-se que o Sistema de Recomendação de Receitas Saudáveis – Tudo Saudável cumpriu a proposta empregada inicialmente oferecendo um sistema simples e prático com a finalidade principal de auxiliar as pessoas de forma ativa na busca por receitas menos calóricas.

8. Cronograma

A seguir é apresentado o cronograma que foi cumprido pela equipe para a inicialização e finalização do projeto nos prazos estabelecidos.

TUDO SAUDÁVEL

Atribuição	Responsáveis	Duração	Início	Conclusão	Progresso
Formulação do Escopo do Projeto	Equipe	25 dias	01/04/2015	25/04/2015	100%
Diretrizes Iniciais	Equipe	10 dias	26/04/2015	10/05/2015	100%
Coleta das Receitas no Site Tudo Gostoso	Equipe	48 dias	11/05/2015	26/06/2015	100%
Mineração dos Ingredientes nas Receitas	Equipe	30 dias	25/06/2015	25/07/2015	100%
Desambiguação dos Ingredientes - I etapa	Equipe	48 dias	20/07/2015	20/08/2015	100%
Relatório Parcial	Equipe	10 dias	27/08/2015	09/09/2015	100%
Desambiguação dos Ingredientes - II etapa	Equipe	7 dias	10/09/2015	18/09/2015	100%
Pesquisa dos Níveis Calóricos	Equipe	7 dias	14/09/2015	22/09/2015	100%
Definição das Substituições	Equipe	15 dias	26/09/2015	15/10/2015	100%
Povoamento do Banco de Dados	Equipe	7 dias	08/10/2015	16/10/2015	100%
Desenvolvimento da Interface Web	Equipe	09 dias	16/10/2015	28/10/2015	100%
Associação ao Banco	Equipe	15 dias	14/09/2015	02/10/2015	100%
Associação à Interface Web	Equipe	10 dias	01/10/2015	14/10/2015	100%
Execução dos Testes de Software	Equipe	7 dias	09/10/2015	19/10/2015	100%
Correção de Falhas	Equipe	5 dias	09/10/2015	19/10/2015	100%
Relatório Final	Equipe	10 dias	22/10/2015	04/11/2015	100%

9. Referências

BULGER B., GREENSPAN J., Aplicações de banco de dados *MySQL / PHP*. 2001.

Livro.

COITINHO D., COUTINHO W F., MONTEIRO J. B., SICHIERI, R., Recomendações de Alimentação e Nutrição Saudável para a População Brasileira Parágrafo Um. 1999.

Artigo.

FERREIRA. Willyan M., SILVA. Ana Paula C., BENEVENUTO. Fabrício, MERSCHMANN. Luiz H. C.; “Comer, Comentar e Compartilhar: Análise de uma Rede de Ingredientes e Receitas.” 2013. Artigo.

FIGUEIRA. Fernando M. Filho, GEUS. Paulo Lício, ALBUQUERQUE. João Porto; “Sistemas de Recomendação e Interação na *Web Social*.” 2008. Artigo.

KOFLER M., O que é o *MySQL*?. 2001. Livro.

LEAVITT N., "Will NOSQL Databases Live Up to Their Promise?" 2010. *Computer*, vol. 43, no. 2, page 12-14.

LIDDY, E. D. “Text Mining. *Bulletin of the the American society for Information Science*, October/November, 2000.” Artigo.

NEWMAN, M. “Assortative mixing in networks. *Physical Review E*, 89(20):208701, 2002.” Artigo.

OMS (Organização Mundial de Saúde), Relatório de uma consulta FAO / OMS. Joint Preparação e utilização de orientações dietéticas baseadas nos alimentos. Genebra, 1998.

OMS (Organização Mundial da Saúde), Relatório de um grupo de consulta sobre a obesidade. Prevenir e gerir a epidemia global de Genebra, 1998.

SICHIERI R, VEIGA G. V., Obesidade na adolescência. In: Juventude, Saúde e Desenvolvimento. Ministério da Saúde, Brasil, 1999. pp: 269-75.

STEVENS J., CAI J., ELSIE R. P., WILANSOM D. F., THUN M. J., MADEIRA J. L., O efeito da idade sobre a associação entre o índice de massa corporal e mortalidade. N Eng J Med 1998.

SUEHRING S., *MySQL™ Bible*. 2002. Livro.

TAUBES G., Como as taxas de obesidade a subir, os especialistas se esforçam para explicar o porquê. 1998. Ciência.

VIGITEL 2014: “Obesidade estabiliza no Brasil, mas excesso de peso aumenta”, disponível em <<http://portalsaude.saude.gov.br/index.php/o-ministerio/principal/secretarias/svs/noticias-svs/17455-obesidade-estabiliza-no-brasil-mas-excesso-de-peso-aumenta>> Acessado em: 04/06/2015 às 19:45.

WILLETT W. C., A perda de peso em idosos: causas ou efeitos. Am J Clin Nutr 1997;.

YANK K., Construa seu próprio banco de dados do site conduzido Utilizando PHP e MySQL. 2004. Livro.

10. Apêndices

10.1. Apêndice A: Script de Coleta

Neste apêndice, é apresentado o script elaborado para a coleta dos ingredientes de cada receita. São esses ingredientes que compõem a base de dados do Sistema. A linguagem utilizada foi Python devido a sua capacidade de tratar grande volume de dados.

```
#-*- coding:iso-8859-1 -*-
import urllib
import urllib2
import string
import sys
import re
from bs4 import BeautifulSoup
from unicodedata import normalize

def remove_acentos(st):
    st = normalize('NFKD',st.decode('utf-8')).encode('ascii', 'ignore')
    return (st)
```

```

def ler_receita(nomereceita):

    receitas = str(nomereceita)
    exp = re.compile(r'<.*?>')
    receitas = exp.sub("", receitas)
    receitas = remove_acentos(exp.sub("", receitas))
    receitas = receitas.strip()
    return receitas

def ler_votos(numvotos):
    receitas = str(numvotos)
    exp = re.compile(r'<.*?>')
    receitas = remove_acentos(exp.sub("", receitas))
    receitas = receitas.strip()
    return receitas

def ler_comentarios(comentarios):

    receitas = str(comentarios)
    exp = re.compile(r'<.*?>')
    receitas = remove_acentos(exp.sub("", receitas))
    receitas = receitas.strip()
    return receitas

def ler_ingrediente(link):
    receitas = str(link)
    exp = re.compile(r'<.*?>')
    receitas = remove_acentos(exp.sub("", receitas))
    receitas = receitas.strip()
    return receitas

def ler_rating(rating):
    receitas = str(rating)
    exp = re.compile(r'<.*?>')
    receitas = remove_acentos(exp.sub("", receitas))
    receitas = receitas.strip()
    receitas = receitas[0]+receitas[1]+receitas[2]
    return receitas

cont = 10901
categoria = "/categorias/1037-doces-e-sobremesas-1091.html"
arq = open('C:/Users/Usuario/Desktop/temp/Base/Doces/receitas1.txt', 'w')
arq1 = open('C:/Users/Usuario/Desktop/temp/Base/Doces/ing_receitas1.txt', 'w')

while True:
    linkp = "http://tudogostoso.uol.com.br"
    response = urllib2.urlopen(linkp+categoria)
    page_source = response.read

```

```

soup = BeautifulSoup(response)
try:
    linkspagest = soup.find('ul', {'class': 'clearfix'})
    for link in linkspagest.find_all('a'):
        if 'receita' in str(link):
            if link.get('href').split('/')[1] == 'receita':
                novolink = linkp+link.get('href')
                if novolink != 'http://tudogostoso.uol.com.br/receita/147941-
bolo-gelado.html':
                    #print ("ID RECEITA: "+str(cont)+ "\n")
                    arq.write("ID RECEITA: "+str(cont)+ "\n")
                    arq1.write("ID RECEITA: "+str(cont)+ "\n")
                    #print("Link: " + novolink + "\n")
                    arq.write("Link: " + novolink + "\n")
                    response1 = urllib2.urlopen(novolink)
                    page_source1 = response1.read
                    soup1 = BeautifulSoup(response1)
                    linha = ""
                    arq4 = str(soup1)
                    for x in arq4:
                        linha = linha + x
                    soup1 = BeautifulSoup(linha)

                    nomereceita = soup1.find('span', {'class': 'current'})
                    nomereceita = ler_receita(nomereceita)
                    #print("Receita: " + nomereceita + "\n")
                    arq.write("Receita: " + nomereceita + "\n")

                    numvotos = soup1.find('span', {'class': 'votenum count'})
                    numvotos = ler_votos(numvotos)
                    #print("Numero de Votos: " + numvotos + "\n")
                    arq.write("Numero de Votos: " + numvotos + "\n")

                    rating = soup1.find('div', {'class': 'ratingbox current
average'})

                    rating = ler_rating(rating)
                    #print("Rating: " + rating + "\n")
                    arq.write("Rating: " + rating + "\n")

                    comentarios = soup1.find('a', {'class': 'commnum'})
                    comentarios = ler_comentarios(comentarios)
                    #print("Comentarios: " + comentarios + "\n")
                    arq.write("Comentarios: " + comentarios + "\n")

                    linkspagestotal = soup1.find('div',
{'class': 'recipelist'})
                    for link in linkspagestotal.find_all('span'):
                        ingrediente = ler_ingrediente(link)
                        #print(ingrediente + "\n")

```

```

        arq1.write(ingrediente + "\n")
    #print "###\n"
    arq.write("###\n")
    arq1.write("###\n")
    #print "###\n"
    if(cont==1 or cont%50==0):
        print "coletei a receita "+str(cont)+": "+nomereceita

    cont = cont + 1

    linkspagestotal = soup.find('a', {'class':'next'})
    categoria = linkspagestotal.get('href')
    print categoria
except AttributeError:
    break
arq.close()
arq1.close()

```

10.2. Apêndice B: Script de Identificação

Neste apêndice, é apresentado o script elaborado para a identificação dos ingredientes de cada receita. São esses ingredientes que compõem a base de dados do Sistema. A linguagem utilizada foi Python devido a sua capacidade de tratar grande volume de dados.

```

#-*- coding:iso-8859-1 -*-
import urllib
import urllib2
import string
import sys
import re
from bs4 import BeautifulSoup
from operator import itemgetter
from unicodedata import normalize

arq = open('C:/Users/Usuario/Desktop/Identificação Final/ing.txt','r')
vetor = arq.readlines()
ing_dif = []
q_ing_dif = []
contador = 0
ccc = 0
i = 0
nr = 1
limsup = 9
liminf = 1
qtd = []
while i < limsup:
    qtd.append(0)

```

```

    i = i + 1
print "num de receitas: "+str(nr)+" de 30668"
print "qtd "+str(contador)+": "+str(qtd)+"\n"

ignorar = open('C:/Users/Usuario/Desktop/Identificação
Final/final/ignorar.txt','r')
ignora= ignorar.readlines()
ignora.reverse()
f_ignorar = open('C:/Users/Usuario/Desktop/Identificação
Final/final/frases_ignorar.txt','r')
f_ign= f_ignorar.readlines()
f_ign.reverse()

#print f_ign

j=0
tam = len(f_ign)
while j < tam:
    aux = f_ign[j]
    aux = aux.split("\n")
    aux = aux[0]
    f_ign[j] = aux
    j = j + 1

j=0
tam = len(ignora)
while j < tam:
    aux = ignora[j]
    aux = aux.split("\n")
    aux = aux[0]
    ignora[j] = aux
    j = j + 1

for linha in vetor:
    if not ("ID RECEITA" in linha):
        if not ("#" in linha):
            linha = linha.split("\n")
            linha = linha[0].lower()
            ll = linha
            linha = linha.split(" ")
            aux = ""
            cont=0
            for naux in linha:
                if not (naux in ignora):
                    if cont > 0:
                        aux = aux + " "+naux
                    else:

```

```

        aux = naux
        cont = cont + 1
linha = aux

if not (linha in f_ign):
    if linha != "":

        #print linha
        achei = 0
        i = limsup
        lim = liminf-1
        while i > lim:
            if achei==1:
                break
            arq1 = open('C:/Users/Usuario/Desktop/Identificação
Final/final/lista_ing_'+str(i)+'.txt', 'r')
            ing = arq1.readlines()
            ing.reverse()
            for x in ing:
                x = x.split("\n")
                x = x[0]

            if x.lower() in linha:
                #print linha+" -> achei na lista de "+str(i)+"
termos: "+ x+"\n\n"

                qtd[i-1] = qtd[i-1] + 1
                contador = contador + 1
                if not (x in ing_dif):
                    ing_dif.append(x)
                    q_ing_dif.append(1)
                else:
                    q_ing_dif[ing_dif.index(x)] =
q_ing_dif[ing_dif.index(x)] + 1

                #if contador%5000 == 0:
                #    print "num de receitas: "+str(nr)+" de 30668"
                #    print "qtd "+str(contador)+": "+str(qtd)+"\n"
                achei = 1
                break
            arq1.close()
            i = i - 1
        if(i==lim and achei==0):
            print ll
            print linha+" -> nao achei\n\n"
    else:
        nr = nr + 1
print "FIM"
print "num de receitas: "+str(nr)+" de 30668"
print "qtd "+str(contador)+": "+str(qtd)+"\n"

```

```

print "foram identificados "+str(contador)+" sendo "+str(len(ing_dif))+
diferentes"
print "media de "+str(float(contador)/30668)+" ingredientes por receita"

teste = []
i = 0
for x in ing_dif:
    teste.append([q_ing_dif[i],x])
    i = i + 1

#teste.sort(0)
teste = sorted(teste, key=itemgetter(0))
for x in teste:
    print x

arq.close()

```

10.3. Apêndice C: Descrição Caso de Uso

Neste apêndice, é apresentado a descrição do Caso de Uso.

Nome do Caso de Uso	
Realizar Login	
Ator Primário	
Usuário	
Atores Secundários	
Não contém	
Resumo	
Este caso de uso descreve as etapas percorridas por um usuário para realizar o login de acesso a sua conta no Sistema.	
Pré-condições	
O usuário deve já estar cadastrado no Sistema.	
Pós-condições	
Após o acesso a conta, usuário será redirecionado à tela principal do Sistema e será disponibilizado novos recursos.	
Fluxo Principal	
Ações do Ator	Ações do Sistema
1 – Informar dados necessários para login.	
	2 – Verificar e validar informações oferecidas pelo ator. 3 – Consultar usuário por código de cadastro.
	5 – Redirecionar para tela principal do Sistema.

Restrições e Validações	
Para login, o usuário deverá entrar com o user e com uma senha de, no mínimo, quatro caracteres.	
Fluxo Alternativo I – Usuário não cadastrado	
Ações do Ator	Ações do Sistema
1 – Informar dados para login.	2 – Verificar e validar dados inseridos. 3 – Informar ao usuário que é necessário a realização de cadastro. 4 – Emitir erro no Sistema.
Fluxo de Exceção I – Falha no acesso da conta	
Ações do Ator	Ações do Sistema
1 – Informar dados para login.	2 – Verificar e validar dados inseridos. 3 – Informar ao ator que não foi possível realizar acesso na conta do Sistema. 4 – Emitir erro no Sistema.

Nome do Caso de Uso	
Gerenciar Receitas	
Ator Primário	
Usuário	
Atores Secundários	
Não contém	
Resumo	
Este caso de uso descreve as etapas percorridas por um usuário para cadastrar, alterar ou excluir uma receita no Sistema.	
Pré-condições	
O usuário deve estar logado em sua conta de acesso do Sistema.	
Pós-condições	
Após a inserção da receita, o usuário será redirecionado para a tela de confirmação da inclusão da receita.	
Fluxo Principal	
Ações do Ator	Ações do Sistema
1 – Realizar login. 2 – Informar dados referentes a receita.	
	3 – Verificar e validar dados inseridos. 4 – Cadastrar receita. 5 – Redirecionar para tela de confirmação do Sistema.
Restrições e Validações	
Para inserção de uma nova receita, deve-se ocorrer a verificação da existência de uma receita igual. Já para manipulação de dados nas receitas, deve-se confirmar a existência da mesma.	
Fluxo Alternativo I – Manipulação de receitas	
Ações do Ator	Ações do Sistema
1 – Informar dados. 2 – Informar ação desejada:	

visualização, inserção, alteração ou exclusão de receita.	3 – Verificar e validar dados inseridos. 4 – Realizar ação desejada pelo administrador. 5 – Redirecionar para tela de confirmação do Sistema.
Fluxo de Exceção I – Falha ao manipular receitas	
Ações do Ator	Ações do Sistema
1 – Informar dados. 2 – Informar ação desejada: visualização, inserção, alteração ou exclusão de receita.	3 – Verificar e validar dados inseridos. 4 – Informar ao usuário que não foi possível realizar a manipulação da receita no Sistema. 5 – Emitir erro no Sistema.

Nome do Caso de Uso	
Visualizar receitas	
Ator Primário	
Usuário	
Atores Secundários	
Não contém	
Resumo	
Este caso de uso descreve as etapas percorridas por um usuário para visualizar as receitas já existentes no Sistema.	
Pré-condições	
Não contém	
Pós-condições	
Após a visualização das receitas, o usuário será redirecionado a tela principal.	
Fluxo Principal	
Ações do Ator	Ações do Sistema
1 – Informar que deseja visualizar receitas.	2 – Verificar e validar dados inseridos. 3 – Visualizar receitas. 4 – Redirecionar para tela de visualização de receitas.
Restrições e Validações	
Para que haja a visualização de receitas, é preciso verificar a existência das mesmas.	
Fluxo de Exceção I – Falha ao visualizar receitas	
Ações do Ator	Ações do Sistema
1 – Informar dados. 2 – Informar ação de visualizar receitas.	3 – Verificar e validar dados inseridos. 4 – Informar ao usuário que não foi possível realizar a visualização da receita. 5 – Emitir erro no Sistema.

Nome do Caso de Uso	
Avaliar recomendação	
Ator Primário	
Usuário	
Atores Secundários	
Não contém	
Resumo	
Este caso de uso descreve as etapas percorridas por um usuário para avaliar uma recomendação de ingrediente do Sistema.	
Pré-condições	
O usuário já deve estar logado em sua conta de acesso do Sistema.	
Pós-condições	
Após avaliar recomendação, o usuário seja redirecionado a tela principal do Sistema.	
Fluxo Principal	
Ações do Ator	Ações do Sistema
1 – Informar dados necessários para login.	
2 – Informar dados para avaliação.	3 – Verificar e validar informações oferecidas pelo ator. 4 – Realizar avaliação.
	5 – Redirecionar para tela principal do Sistema.
Fluxo de Exceção I – Falha ao avaliar	
Ações do Ator	Ações do Sistema
1 – Informar dados para avaliação.	2 – Verificar e validar dados inseridos. 3 – Informar ao usuário que não foi possível avaliar a recomendação no Sistema. 4 – Emitir erro no Sistema.

Nome do Caso de Uso	
Realizar comentário	
Ator Primário	
Usuário	
Atores Secundários	
Não contém	
Resumo	
Este caso de uso descreve as etapas percorridas por um usuário para comentar uma recomendação de ingrediente no Sistema.	
Pré-condições	
O usuário já deve estar logado em sua conta de acesso do Sistema.	
Pós-condições	
Após comentar uma recomendação, o usuário seja redirecionado a tela principal do Sistema.	

Fluxo Principal	
Ações do Ator	Ações do Sistema
1 – Informar dados necessários para login.	
2 – Informar dados para comentário.	3 – Verificar e validar informações oferecidas pelo ator. 4 – Manter comentário.
	5 – Redirecionar para tela principal do Sistema.
Restrições e Validações	
Para comentar uma recomendação, o usuário deverá proporcionar um comentário de no máximo 150 caracteres.	
Fluxo de Exceção I – Falha ao comentar	
Ações do Ator	Ações do Sistema
1 – Informar dados para comentário.	2 – Verificar e validar dados inseridos. 3 – Informar ao usuário que não foi possível inserir comentário à recomendação no Sistema. 4 – Emitir erro no Sistema.