

Automatic Recipe Ingredient Substitution Based on Text Mining and Data Clustering Approaches

Luciano D. S. Pacífico¹, Larissa F. S. Britto^{2,3}, Teresa B. Ludermir²

¹ Departamento de Computação (DC), Universidade Federal Rural de Pernambuco (UFRPE), Brazil
luciano.pacifico@ufrpe.br

² Centro de Informática (CIn), Universidade de Federal de Pernambuco (UFPE), Brazil
{lfsb, tbl}@cin.ufpe.br

³ Centro de Pesquisa e Desenvolvimento em Telecomunicações (CPQD), Brazil

Abstract. With the advance of Information Technologies, recipe sharing websites have become very common, and people can use such systems in an attempt to find a recipe that fits both their desires and nutritional needs. But sometimes ingredients from a given recipe list may not be available, what may require some adaptations by replacing the missing ingredients. In this work, a data-driven approach is employed to develop a new recipe generation system, that recommends substitute ingredients to adapt recipes into a target domain. The proposed system, which is based on Text Mining and Data Clustering techniques, is evaluated by means of a qualitative analysis, showing promising results.

CCS Concepts: • **Applied computing** → **Consumer health**; • **Information systems** → **Recommender systems**; *Collaborative filtering*.

Keywords: Data Clustering, Natural Language Processing, Recommendation Systems, Text Mining

1. INTRODUÇÃO

O compartilhamento de receitas, outrora realizado através de cadernos, livros e revistas especializadas, mas que atualmente é feito via meios de comunicação digitais, como websites e comunidades online, tem facilitado a busca por receitas de diferentes culinárias do mundo, auxiliando usuários com níveis de habilidades variados (desde os totalmente inexperientes, aos chefs profissionais) a encontrarem refeições completas das mais diversas culturas. A popularidade da Internet possibilita, nos dias atuais, que qualquer pessoa possa aprender a preparar pratos que lhes satisfaçam sem sair de sua casa.

Porém, em decorrência da enorme sobrecarga de dados (receitas) presentes em páginas especializadas, muitas vezes os usuários desses sites têm dificuldades em encontrar um prato específico, que lhes seja agradável, ou ainda, refeições completas que contemplem tanto suas necessidades nutricionais, quanto seus gostos pessoais. Nesse contexto, diversos Sistemas de Recomendação de Receitas têm sido propostos como uma tentativa de auxiliar usuários no encontro da refeição ideal para um dado momento do dia [Pecune et al. 2020]. Mas nem sempre todos os ingredientes da lista das receitas sugeridas pelos sistemas de recomendação estão à disposição, havendo necessidade de realização de adequações e substituições [Ławrynowicz et al. 2022; Shino et al. 2016]. O problema é ainda mais grave quando o usuário possui algum tipo de restrição alimentar, uma vez que os repositórios web possuem receitas predominantemente voltadas ao público geral (sem restrições), o que torna a necessidade da substituição de ingredientes nocivos imprescindível para adequar essas receitas às necessidades desses usuários [Pacífico et al. 2021; Ooi et al. 2015].

Neste trabalho, uma abordagem de geração de receitas pela substituição de um ingrediente único,

Copyright©2022 Permission to copy without fee all or part of the material printed in KDMiLe is granted provided that the copies are not made or distributed for commercial advantage, and that notice is given that copying is by permission of the Sociedade Brasileira de Computação.

baseada em uma metodologia voltada aos dados, é proposta. Técnicas de Mineração de Textos e Análise de Agrupamentos são empregadas para avaliar a aceitação e adequação de ingredientes substitutos candidatos, considerando o contexto das receitas a serem convertidas e a relação entre ingredientes. O trabalho está organizado da seguinte forma: a Seção 2 apresenta a metodologia do sistema de geração de receitas proposto, seguido pela análise experimental (Seção 3). Por fim, as conclusões e linhas de trabalhos futuros serão apresentadas (Seção 4).

2. METODOLOGIA

A metodologia adotada no presente trabalho será detalhada nesta seção. Uma visão geral do sistema de geração de receitas proposto é ilustrada na Fig. 1. O sistema proposto é dividido em duas etapas básicas, que são a aquisição dos dados e o sistema de filtragem, descritos em detalhes na sequência (Seção 2.1 e Seção 2.2, respectivamente).

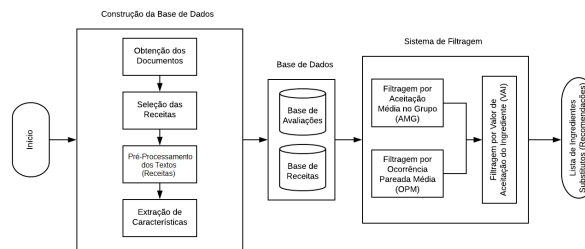


Fig. 1. Sistema de geração de receitas proposto.

2.1 Base de Dados

Como o sistema de recomendação de ingredientes substitutos proposto é baseado em uma abordagem voltada aos dados, a seleção de uma base de receitas adequada é fundamental para o seu bom desempenho. Nesse intuito, o presente trabalho fará uso da base de dados apresentada em [Majumder et al. 2019], extraída da página Food.com¹, que apresenta as receitas como **documentos textuais** no idioma inglês. Para cada receita estão disponíveis informações como lista de ingredientes, modos de preparo, valores nutricionais por categoria, comentários e avaliações dos usuários, além de uma categorização feita por seus usuários através de um sistema de *tags*. Como o intuito do trabalho atual é a recomendação automática de ingredientes substitutos, apenas as informações relacionadas à **lista de ingredientes**, às *tags* das receitas e às **avaliações dos usuários para cada uma dessas receitas** (representadas por um sistema de *ranking*, que varia entre a nota 0 - pior avaliação - à nota 5 - melhor avaliação) serão levadas em consideração. Embora o sistema de recomendação de ingrediente substituto seja implementado como uma abordagem automática de propósito geral, no presente trabalho, estamos interessados na **geração de receitas que sejam seguras às restrições alimentares dos usuários** (como para o caso de alérgicos, ou usuários com algum outro tipo de restrição alimentar), pela substituição de um único ingrediente que causaria tal restrição [Pacífico et al. 2021; Ooi et al. 2015]. Para isso, três categorias de restrições alimentares foram escolhidas: receitas seguras para **Diabéticos**, receitas **Sem Glúten** e receitas **Sem Lactose**. A tarefa final executada pelo sistema será a conversão de uma receita sem restrição conhecida, *a priori*, para algum dos três domínios de restrições alimentares selecionados, pela substituição de um único ingrediente que infringiria à restrição em questão.

A base de dados original já contém algum grau de pré-processamento [Majumder et al. 2019]. Mesmo assim, para o presente trabalho, optou-se pela realização de uma nova etapa de pré-processamento

¹<https://www.kaggle.com/shuyangli94/food-com-recipes-and-user-interactions>

Table I. Estatísticas da Base de Dados.

Categoria das Receitas	Diabéticos	Sem Glúten	Sem Lactose
Tag(s) Originais	'diabetic'	'gluten-free'	'lactose', 'dairy-free', 'vegan'
Número de Receitas	6458	5743	13129
Número de Ingredientes (sem Pré-Processamento)	4487	3793	5205
Número de Ingredientes (após Pré-Processamento)	4155	3492	4766
Número Médio de Ingredientes por Receita	8.8587	8.5195	8.89793
Número Total de Avaliações	30786	32226	66580
Número Médio de Avaliações por Receita	4.7671	5.6113	5.0712
Média Acumulada de Aceitação dos Usuários por Receita	20.7951	24.5552	22.2665
Média Absoluta de Aceitação dos Usuários por Receita	4.3622	4.3760	4.3907

dos dados textuais, através de técnicas de Processamento de Linguagem Natural (*Natural Language Processing*, ou NLP), como forma de promover uma melhor limpeza de tais dados, assim como a eliminação de redundâncias e características irrelevantes encontradas na base original. O novo pré-processamento foi realizado de acordo com as seguintes etapas:

- Conversão para Minúscula:** Todas as letras maiúsculas encontradas nos documentos foram convertidas para minúsculas, de modo a evitar que conceitos semanticamente idênticos para os propósitos da aplicação, como ‘egg’ e ‘Egg’ (‘ovo’), sejam reconhecidos como termos diferentes.
- Conversão de Caracteres Especiais:** Caracteres especiais foram convertidos para os caracteres equivalentes na tabela ASCII (por exemplo, ‘ç’ é convertido em ‘c’).
- Remoção de URLs e Pontuações:** Para os propósitos da aplicação atual, pontuação e URL não são relevantes, uma vez que apenas a ocorrência dos ingredientes nos textos é de interesse.
- Remoção de Stop Words:** As *stop words* são termos não informativos presentes frequentemente em documentos textuais, como conjunções e preposições.
- Lematização:** Lematização é o processo de agrupamento de representações diferentes de um mesmo termo, de modo a possibilitar a identificação dessas formas como um único conceito. Na base de dados original, termos como ‘egg’ e ‘eggs’ são reconhecidos como conceitos diferentes, mas para os propósitos do trabalho atual, ambos se referem ao mesmo conceito/ingrediente (‘ovo’).

Na abordagem atualmente proposta, o método de **extração de características *Bag-of-Words Binária*** é empregado para a representação dos documentos textuais (ou seja, das receitas) que serão levados em consideração pelo sistema de filtragem e geração de receitas. Na representação por *Bag-of-Words Binária*, o conjunto de documentos será convertido em uma **Matriz de Termos do Documento** (*Document-Term Matrix*, ou DTM), na qual cada linha representará um documento (ou seja, uma **receita**), e cada coluna representa um termo (ou seja, um **ingrediente**). Cada posição na DTM indicará se um dado ingrediente faz parte (indicado por ‘1’) ou não (indicado por ‘0’) da lista de ingredientes da receita em questão. Esta abordagem se mostra bastante adequada à proposta atual, uma vez que desejamos avaliar apenas se um ingrediente faz parte ou não da lista de uma dada receita. A Tabela I apresenta algumas estatísticas para a base de dados final, obtida após o pré-processamento dos textos e extração de características. A quantidade de ingredientes é apresentada antes e após a redução dos dados pelo pré-processamento. A *Média Acumulada de Aceitação dos Usuários por Receita* refere-se à soma total dos *ranks* de todas as avaliações recebidas pelas receitas da categoria, dividida pelo número total de receitas nessa categoria. A *Média Absoluta de Aceitação dos Usuários por Receita* é calculada como a soma total dos *ranks* de todas as avaliações recebidas pelas receitas da categoria, dividida pelo número total de avaliações de receitas na categoria em questão. A Figura 2 apresenta os ingredientes mais frequentes nas categorias de receitas selecionadas.

2.2 Sistema de Filtragem

Após a aquisição das bases de dados (receitas, representadas por suas listas de ingredientes, e avaliações, representadas pelas qualificações numéricas), um sistema de filtragem é proposto para a recomendação dos ingredientes substitutos para a geração das novas receitas. O sistema de filtragem funcionará da seguinte forma: dada uma receita base $\mathbf{r}_i \in \mathbf{R}_s$ (domínio de receitas sem restrições, a

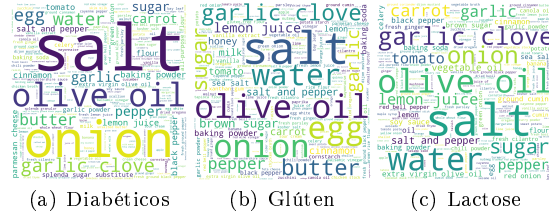


Fig. 2. Ingredientes mais frequentes nas categorias de receitas selecionadas.

priori), e um ingrediente alvo $x_t \in \mathbf{r}_i$ a ser substituído, deseja-se encontrar os melhores ingredientes candidatos à substituição x_c de x_t no domínio destino \mathbf{R}_t , gerando, desta forma, uma nova receita \mathbf{r}_t que seja adequada às restrições desse domínio destino.

Nesse intuito, dois filtros serão usados: um filtro baseado no contexto das receitas a serem convertidas, e um filtro que leva em consideração as relações dos ingredientes candidatos x_c à substituição do x_t com os demais ingredientes em \mathbf{r}_i ; por fim, ambos os filtros serão combinados para a geração do *rank* final de cada x_c .

O primeiro filtro do sistema, o filtro de **Aceitação Média no Grupo (AMG)**, consiste na avaliação da aceitação dos ingredientes candidatos no contexto da receita base \mathbf{r}_i . Primeiramente, as receitas do domínio destino \mathbf{R}_t são separadas em grupos, de acordo com suas listas de ingredientes (representadas pela DTM), pelo uso de um algoritmo de agrupamento. Na abordagem atual, o algoritmo de agrupamento selecionado foi o K-Means [MacQueen et al. 1967], por sua simplicidade e velocidade de execução. Após a obtenção dos grupos, o contexto da receita base \mathbf{r}_i é obtido pela determinação de seu grupo de receitas \mathbf{G}_i no \mathbf{R}_t . Uma vez que o contexto da \mathbf{r}_i é determinado, é realizada uma etapa de cálculo da aceitação média de cada ingrediente candidato x_c nesse contexto. Para isso, as avaliações dos usuários $\mathbf{E}_{\mathbf{G}_i}$, representadas por seus valores de *rank*, realizadas em todas as receitas em \mathbf{G}_i serão levadas em consideração. A aceitação do ingrediente candidato x_c em relação à receita $\mathbf{r}_j \in \mathbf{G}_i$ é dada por (eq. (1)):

$$V_{ac}(x_c, \mathbf{r}_j) = \begin{cases} \frac{\sum_{eval \in \mathbf{E}_{\mathbf{r}_j}} rank(eval, \mathbf{r}_j)}{\sum_{\mathbf{r}_j \in \mathbf{G}_i} \sum_{eval \in \mathbf{E}_{\mathbf{r}_j}} rank(eval, \mathbf{r}_j)}, & \text{se } x_c \in \mathbf{r}_j \\ 0, & \text{caso contrário} \end{cases} \quad (1)$$

onde *eval* refere-se à avaliação de um usuário para uma receita \mathbf{r}_j , $\mathbf{E}_{\mathbf{r}_j}$ refere-se ao conjunto de avaliações de usuários para uma receita \mathbf{r}_j , e *rank*(•) retorna o *rank* numérico de uma avaliação *eval* da receita \mathbf{r}_j . Por fim, o valor da Aceitação Média no Grupo para x_c é dada por (eq. (2)):

$$AMG(x_c) = \sum_{\mathbf{r}_j \in \mathbf{G}_i} V_{ac}(x_c, \mathbf{r}_j) \quad (2)$$

O segundo filtro do sistema avalia a relação dos ingredientes candidatos x_c com os demais ingredientes $x_i \in \mathbf{r}_i$. Essa avaliação faz-se necessária pelo fato de que, em uma abordagem puramente voltada aos dados, uma das formas de determinação da adequação da combinação de ingredientes é obtida pela verificação da ocorrência pareada dos ingredientes nas receitas da base de dados [Nirmal et al. 2018]. Uma alta ocorrência de determinados pares de ingredientes em diferentes receitas é um bom indicativo de que essa combinação é considerada adequada e agradável pelos usuários, além de representar uma combinação segura (não tóxica). Considere a lista de ingredientes total \mathbf{I}_t no domínio destino \mathbf{R}_t . Para cada ingrediente candidato $x_c \in \mathbf{I}_t$, o filtro de **Ocorrência Pareada Média (OPM)** visa analisar a relação de cada x_c com os demais ingredientes x_i ($\{x_i \in \mathbf{r}_i | x_i \neq x_t\}$) da receita base \mathbf{r}_i . Para isso, faz-se necessário o cálculo da Matriz de Ocorrência Pareada $\mathbf{M}_p(x_c, x_i)$ entre cada ingrediente $\{x_c \in \mathbf{I}_t | x_c \notin \mathbf{r}_i\}$ e todos os ingredientes $x_i \in \mathbf{r}_i (x_i \neq x_t)$, considerando todas as receitas no domínio destino \mathbf{R}_t . O OPM é então calculado como (eq. (3)):

$$OPM(x_c) = \sum_{x_i \in \mathbf{r}_i (x_i \neq x_t)} \mathbf{M}_p(x_c, x_i) / (|\mathbf{r}_i| - 1) \quad (3)$$

Table II. Receitas selecionadas para a avaliação experimental.

Receita	Número de Ingredientes	Ingredientes
Quick Pound Cake	5	flour, egg, sugar, butter, almond extract
Simple Fudge Brownies	6	all-purpose flour, butter, cocoa, egg, sugar, vanilla
Wheat Germ and Banana Muffins	10	all-purpose flour, baking powder, banana, egg, milk, salt, sugar, toasted wheat germ, vanilla extract, vegetable oil
Whole Wheat Corn Bread	8	baking powder, cornmeal, egg, milk, salt, sugar, vegetable oil, whole wheat flour

onde $|\mathbf{r}_i|$ indica o número total de ingredientes na receita base \mathbf{r}_i . Como apenas os ingredientes no domínio destino são levados em consideração no cálculo do OPM, há uma garantia de que se as receitas da base de dados respeitarem às restrições nesse domínio destino, não haverá recomendação de ingredientes inadequados.

Por fim, os resultados para o AMG e para o OPM são combinados para a formação do **Valor de Aceitação do Ingrediente (VAI)**, que representa o escore final que será levado em consideração para a recomendação dos x_c na geração das novas receitas \mathbf{r}_t . Para isso, os seguintes passos são adotados:

- (1) Para cada ingrediente candidato $x_c \in \mathbf{I}_t$, seus valores de $AMG(x_c)$ e $OPM(x_c)$ são calculados.
- (2) Para cada um dos filtros (AMG e OPM) individualmente, os ingredientes são ordenados do melhor (maiores escores) para o pior (escore mais baixo).
- (3) A cada ingrediente é atribuído um novo valor para os filtros, resultando nos valores $AMG_n(x_c)$ e $OPM_n(x_c)$, de acordo com sua posição no vetor ordenado do filtro em questão, sendo os maiores valores atribuídos aos melhores ingredientes candidatos (eq. (4)).

$$AMG_n(x_c) = |I_t| - pos_{AMG(x_c)}, \quad OPM_n(x_c) = |I_t| - pos_{OPM(x_c)} \quad (4)$$

onde $|I_t|$ refere-se ao número total de ingredientes no domínio alvo, $pos_{AMG(x_c)}$ e $pos_{OPM(x_c)}$ referem-se à posição do ingrediente candidato x_c no vetor ordenado para os filtros $AMG(x_c)$ e $OPM(x_c)$, respectivamente.

- (4) Por fim, os novos valores $AMG_n(x_c)$ e $OPM_n(x_c)$ são combinados, de acordo com a eq. (5):

$$VAI(x_c) = p_1 \cdot AMG_n(x_c) + p_2 \cdot OPM_n(x_c) \quad (5)$$

onde p_1 e p_2 são pesos configuráveis atribuídos aos filtros propostos.

3. ANÁLISE DOS RESULTADOS

Nesta seção, o sistema de geração de receitas pela substituição de ingredientes proposto será avaliado através de uma análise qualitativa. Neste intuito, duas questões de pesquisa principais são elaboradas:

- (1) O sistema é capaz de encontrar bons ingredientes candidatos à substituição automaticamente, levando em consideração a receita base escolhida?
- (2) Os filtros propostos e sua combinação se mostraram adequados?

A avaliação do sistema proposto levará em consideração quatro receitas exemplo, selecionadas do conjunto de receitas sem restrições (Domínio Base). De modo semelhante, um ingrediente alvo de cada uma das receitas será escolhido para exemplificar a execução do sistema proposto, sendo esse ingrediente um ingrediente restrito no Domínio Destino em análise. A lista de ingredientes de cada uma das receitas selecionadas é apresentada na Tabela II.

A análise qualitativa proposta levará em consideração a avaliação dos dez melhores ingredientes candidatos à substituição do ingrediente alvo (análise **Top-10**) [Ooi et al. 2015; Shino et al. 2016; Nirmal et al. 2018], obtidos pelos filtros AMG e OPM, assim como o resultado **Top-10** para a combinação desses filtros (representada pelo filtro VAI). Essa análise é realizada com o auxílio da consulta a websites especializados. De acordo com essas consultas, as seguintes considerações foram realizadas:

- Substituições adequadas em ‘Diabéticos’:** para todas as receitas analisadas, o ingrediente alvo selecionado foi *sugar* (*açúcar*). Para o tipo das receitas selecionadas, boas substituições para o açúcar seriam os ingredientes *vanilla*, *vanilla extract*, *cinnamon*, adoçantes à base de *stevia*, adoçantes à base de sucralose, algumas frutas frescas, etc².
- Substituições adequadas para ‘Sem Glúten’:** os ingredientes *all-purpose flour*, *flour*, e *whole wheat flour* são selecionados como ingredientes alvo em suas receitas originais. Para os ingredientes selecionados, substituições adequadas seriam representadas por ingredientes como *xathan gum*, *rice flour*, *coconut flour*, *tapioca flour*, *potato flour*, *cornstarch* e *potato starch*, por exemplo³.
- Substituições adequadas para ‘Sem Lactose’:** para o ingrediente alvo *butter* (*manteiga*), podemos considerar que qualquer tipo de óleo (como os ingredientes *vegetable oil*, *olive oil*, dentre outros, que são óleos de origem vegetal), além de alguns outros ingredientes, como alguns tipos de purê de frutas, são boas substituições no contexto das receitas *Quick Pound Cake* e *Simple Fudge Brownies*⁴; Já para o caso do ingrediente alvo *milk* (*leite*), no contexto das receitas *Wheat Germ and Banana Muffins* e *Whole Wheat Corn Bread*, podemos considerar que os ingredientes *water*, *soymilk* e outros tipos de leites de origem vegetal são candidatos adequados à substituição no domínio destino escolhido, uma vez que tais ingredientes constam nas listas de boas substituições para o leite em receitas de assados (*baked goods*)^{5,6}.

Os resultados experimentais para o filtro AMG são apresentados na Tabela III. Nos experimentos, o número de agrupamentos finais para o K-Means, escolhido após um processo de tentativa-e-erro, foi $k = 5$, e esse algoritmo executou por 100 iterações. Como os resultados do AMG não estão relacionados ao ingrediente alvo escolhido, e sim, ao contexto da receita base como um todo no Domínio Destino, em alguns casos não foi possível a obtenção de um substituto diretamente adequado ao ingrediente alvo na avaliação *top-10*, o que é um resultado esperado (como para o Domínio Destino ‘Diabéticos’, onde o grupo que continha as receitas *Simple Fudge Brownies*, *Wheat Germ and Banana Muffins* e *Whole Wheat Corn Bread* não encontrou um bom substituto para o ingrediente alvo *sugar*). Porém, para um melhor entendimento do contexto da receita base no domínio alvo, tal filtro mostrou-se bastante eficaz, uma vez que ingredientes indicados para os tipos de preparo avaliados foram priorizados, como o ingrediente *buttermilk*, que embora não seja um ingrediente substituto direto para *sugar*, está bastante associado ao preparo de receitas saudáveis de assados no domínio ‘Diabéticos’.

Já para o filtro OPM (Tabela IV), que leva em consideração o ingrediente alvo em análise, ao menos um ingrediente candidato adequado foi encontrado para cada uma das receitas bases, em cada Domínio Destino.

Após uma avaliação por tentativa-e-erro, os valores dos pesos dos filtros AMG e OPM no cálculo do VAI foram determinados de modo que $p_1 = 0.3$, e $p_2 = 0.7$. Os valores finais para o VAI são apresentados na Tabela V. Como pode ser observado, os resultados finais contiveram ao menos um ingrediente candidato adequado, e nenhum ingrediente que infrija as regras do domínio destino foi recomendado, o que resultaria na geração de receitas seguras nesse domínio. Além disso, para a maioria dos casos, a combinação dos filtros AMG e OPM resultou em recomendações melhores do que as obtidas pelos filtros individualmente. Assim, podemos considerar que o sistema de filtragem proposto é adequado e capaz de gerar boas recomendações levando em conta o contexto das receitas base e os ingredientes a serem substituídos.

²<https://www.diabetesselfmanagement.com/nutrition-exercise/meal-planning/baking-and-cooking-with-sugar-substitutes/>

³<https://www.thekitchn.com/16-gluten-free-flour-alternatives-22943791>

⁴<https://delishably.com/dairy/Substitutes-for-Butter-in-Baking>

⁵<https://blogchef.net/substitute-for-milk-in-cornbread/>

⁶<https://www.sheknows.com/food-and-recipes/articles/1116425/milk-substitutions/>

Table III. Resultados experimentais *Top-10* para o AMG. Substituições adequadas estão em **negrito**.

Receita		Aceitação Média no Grupo (AMG)							
Ingrediente Alvo		Quick Pound Cake		Simple Fudge Brownie		Wheat Germ Banana Muffin		Whole Wheat Corn Bread	
		sugar		sugar		sugar		sugar	
		Candidato	Escore	Candidato	Escore	Candidato	Escore	Candidato	Escore
Diabéticos	olive oil	0.1534		baking powder	0.7577	baking powder	0.7577	baking powder	0.7577
	water	0.1031		baking soda	0.6570	baking soda	0.6570	baking soda	0.6570
	egg	0.1289		egg	0.5273	egg	0.5273	egg	0.5273
	garlic	0.1110		salt	0.4759	salt	0.4759	salt	0.4759
	butter	0.09612		all-purpose flour	0.2783	all-purpose flour	0.2781	all-purpose flour	0.2781
	salt and pepper	0.0781		butter/milk	0.2783	butter/milk	0.2785	butter/milk	0.2785
	cinnamon	0.0775		unsalted butter	0.2630	unsalted butter	0.2630	unsalted butter	0.2630
	parmesan cheese	0.0694		unbleached all-purpose flour	0.2386	unbleached all-purpose flour	0.2386	unbleached all-purpose flour	0.2386
	lemon juice	0.0620		baker salt	0.2368	baker salt	0.2368	baker salt	0.2368
	garlic powder	0.0603		whole wheat flour	0.2189	whole wheat flour	0.2189	whole wheat flour	0.2189
Adequate Substitutes		1		0		0		0	
Receita		Aceitação Média no Grupo (AMG)							
Ingrediente Alvo		Quick Pound Cake		Simple Fudge Brownie		Wheat Germ Banana Muffin		Whole Wheat Corn Bread	
		flour		flour		all-purpose flour		wheat flour	
		Candidato	Escore	Candidato	Escore	Candidato	Escore	Candidato	Escore
Sem Glúten	egg	0.8881		egg	0.8881	egg	0.8881	egg	0.8881
	salt	0.7380		salt	0.7380	salt	0.7380	salt	0.7380
	sugar	0.2000		sugar	0.2000	sugar	0.2000	sugar	0.2000
	baking powder	0.4554		baking powder	0.4554	baking powder	0.4554	baking powder	0.4554
	butter	0.3979		butter	0.3979	butter	0.3979	butter	0.3979
	baking soda	0.3660		baking soda	0.3660	baking soda	0.3660	baking soda	0.3660
	xanthan gum	0.3423		xanthan gum	0.3423	xanthan gum	0.3423	xanthan gum	0.3423
	rice flour	0.2996		rice flour	0.2996	rice flour	0.2996	rice flour	0.2996
	vanilla	0.2372		vanilla	0.2372	vanilla	0.2372	vanilla	0.2372
	potato starch	0.2091		potato starch	0.2091	potato starch	0.2091	potato starch	0.2091
Substitutos Adequados		3		3		3		3	
Receita		Aceitação Média no Grupo (AMG)							
Ingrediente Alvo		Quick Pound Cake		Simple Fudge Brownie		Wheat Germ Banana Muffin		Whole Wheat Corn Bread	
		butter		butter		milk		milk	
		Candidato	Escore	Candidato	Escore	Candidato	Escore	Candidato	Escore
Sem Lactose	water	0.2231		water	0.2231	salt	0.8881	salt	0.8881
	sugar	0.1920		sugar	0.1920	baking soda	0.6135	baking soda	0.6135
	garlic clove	0.1017		garlic clove	0.1017	baking powder	0.5911	baking powder	0.5911
	soy sauce	0.1154		soy sauce	0.1154	sugar	0.5272	sugar	0.5272
	brown sugar	0.0808		brown sugar	0.0808	butter	0.3872	flour	0.3872
	extra virgin olive oil	0.0768		extra virgin olive oil	0.0768	water	0.3087	water	0.3087
	lemon	0.0941		lemon	0.0941	soymilk	0.2907	soymilk	0.2907
	banana	0.0689		banana	0.0689	cinnamon	0.2604	cinnamon	0.2604
	garlic	0.0643		garlic	0.0643	vanilla	0.2551	vanilla	0.2551
	lemon juice	0.0639		lemon juice	0.0639	vegetable oil	0.2476	vegetable oil	0.2476
Substitutos Adequados		1		1		2		2	

Table IV. Resultados experimentais *Top-10* para o OPM. Substituições adequadas estão em **negrito**.

Receita		Ocorrência Parada Média (OPM)							
Ingrediente Alvo		Quick Pound Cake		Simple Fudge Brownie		Wheat Germ Banana Muffin		Whole Wheat Corn Bread	
		sugar		sugar		sugar		sugar	
		Candidato	Escore	Candidato	Escore	Candidato	Escore	Candidato	Escore
Diabéticos	baking powder	0.1100		baking powder	0.1123	baking soda	0.1282	baking soda	0.1407
	salt	0.1018		baking soda	0.1297	whole wheat flour	0.0668	all-purpose flour	0.0982
	baking soda	0.0949		salt	0.0919	flour	0.0665	flour	0.0759
	milk	0.0759		splenda granular	0.0750	cinnamon	0.0640	butter	0.0714
	vanilla	0.0678		cinnamon	0.0699	flour	0.0617	vanilla	0.0692
	cinnamon	0.0594		flour	0.0675	vanilla	0.0606	cinnamon	0.0689
	all-purpose flour	0.0559		milk	0.0659	splenda granular	0.0583	vanilla extract	0.0607
	vanilla extract	0.0514		splenda sugar substitute	0.0573	splenda sugar substitute	0.0496	onion	0.0527
	onion	0.0499		whole wheat flour	0.0551	egg white	0.0469	splenda granular	0.0402
	splenda granular	0.0496		vanilla extract	0.0513	water	0.0439	water	0.0496
Substitutos Adequados		4		4		4		4	
Receita		Ocorrência Parada Média (OPM)							
Ingrediente Alvo		Quick Pound Cake		Simple Fudge Brownie		Wheat Germ Banana Muffin		Whole Wheat Corn Bread	
		flour		flour		all-purpose flour		wheat flour	
		Candidato	Escore	Candidato	Escore	Candidato	Escore	Candidato	Escore
Sem Glúten	salt	0.1323		baking powder	0.1256	baking soda	0.1192	baking soda	0.1084
	baking powder	0.1106		salt	0.1173	xanthan gum	0.0985	xanthan gum	0.1000
	milk	0.0965		baking soda	0.1096	butter	0.0897	butter	0.0951
	baking soda	0.0937		milk	0.0940	vanilla	0.0760	vanilla	0.0772
	xanthan gum	0.0834		xanthan gum	0.0804	cinnamon	0.0724	rice flour	0.0726
	vanilla	0.0768		rice flour	0.0777	rice flour	0.0686	water	0.0724
	water	0.0722		brown sugar	0.0702	water	0.0670	water	0.0666
	rice flour	0.0682		cinnamon	0.0671	gluten-free flour	0.0648	gluten-free flour	0.0641
	vanilla extract	0.0643		water	0.0648	brown sugar	0.0601	cornstarch	0.0580
	cinnamon	0.0622		gluten-free flour	0.0586	potato starch	0.0583	potato starch	0.0580
Substitutos Adequados		2		3		4		5	
Receita		Ocorrência Parada Média (OPM)							
Ingrediente Alvo		Quick Pound Cake		Simple Fudge Brownie		Wheat Germ Banana Muffin		Whole Wheat Corn Bread	
		butter		butter		milk		milk	
		Candidato	Escore	Candidato	Escore	Candidato	Escore	Candidato	Escore
Sem Lactose	baking powder	0.1152		baking soda	0.1209	baking soda	0.1141	baking soda	0.1161
	baking soda	0.1070		baking powder	0.1131	water	0.0650	water	0.0790
	salt	0.0711		salt	0.0611	cinnamon	0.0643	flour	0.0725
	vanilla	0.0656		flour	0.0600	soymilk	0.0642	all-purpose flour	0.0685
	vanilla extract	0.0580		cinnamon	0.0565	flour	0.0603	cinnamon	0.0673
	cinnamon	0.0576		soymilk	0.0541	whole wheat flour	0.0571	vanilla	0.0602
	water	0.0564		whole wheat flour	0.0522	vanilla	0.0535	soymilk	0.0601
	oil	0.0498		water	0.0481	canola oil	0.0453	vanilla extract	0.0591
	soymilk	0.0482		vanilla extract	0.0473	brown sugar	0.0427	onion	0.0542
	brown sugar	0.0411		brown sugar	0.0428	onion	0.0423	canola oil	0.0473
Substitutos Adequados		2		1		2		2	

4. CONCLUSÕES

Neste trabalho, um sistema de geração de receitas pela substituição de ingredientes é proposto, baseado em uma abordagem de Análise de Textos e Agrupamento de Dados. Através de um processo automático de filtragem e de uma abordagem voltada aos dados, o sistema proposto é capaz de oferecer ingredientes que sejam bons candidatos à substituição de um ingrediente alvo, levando em consideração a aceitação desses ingredientes candidatos no contexto da receita, assim como a compatibilidade desses substitutos em relação aos demais ingredientes da receita. O sistema proposto leva em consideração a lista de ingredientes e as avaliações dos usuários recebidas por cada receita da base de dados adotada. Uma análise qualitativa é usada como forma de avaliação da adequabilidade dos *Top-10* melhores ingredientes candidatos à substituição retornados pelo sistema de filtragem proposto. Os resultados experimentais indicam que o sistema proposto é capaz de recomendar ingredientes substitutos adequados às receitas originais, e que não infrinjam as restrições do domínio de receitas destino,

Table V. Resultados *Top-10* finais para o sistema proposto (filtro VAI). Substituições adequadas estão em **negrito**.

Receita		Valor de Avaliação do Ingrediente (VAI)									
Receita		Quick Pound Cake		Simple Pudge Brownie		Wheat Germ Banana Muffin		Whole Wheat Corn Bread			
Ingrediente Alvo		sugar		sugar		sugar		sugar			
		Candidato	Escore	Candidato	Escore	Candidato	Escore	Candidato	Escore	Candidato	Escore
Diabéticos	cinamon	4149.7	4149.7	baking powder	4153.8	baking soda	4154.7	baking soda	4154.7	baking soda	4154.7
	water	4147.7		baking soda	4153.8	whole wheat flour	4153.6	all-purpose flour	4153.1	all-purpose flour	4153.1
	oil	4146.9		oil	4152.7	cinamon	4149.9	flour	4149.1	flour	4149.1
	oil and pepper	4145.8		cinamon	4149.2	oil	4149.4	vanilla	4148.9	vanilla	4148.9
	vanilla	4145.3	4145.3	splenda granular	4147.8	vanilla	4148.2	water	4148.7	water	4148.7
	vine oil	4143.8		flour	4147.9	flour	4145.7	cinamon	4148.5	cinamon	4148.5
	parmesan cheese	4143.1		whole wheat flour	4146.7	splenda granular	4145.7	vanilla extract	4146.9	vanilla extract	4146.9
	vanilla extract	4142.6		splenda sugar substitute	4145.3	splenda sugar substitute	4145.3	splenda granular	4144.3	splenda granular	4144.3
	splenda sugar substitute	4142.0	4142.0	vanilla extract	4144.8	egg white	4144.0	egg white	4142.6	egg white	4142.6
	splenda granular	4136.4	4136.4	oil	4143.3	ground cinnamon	4141.6	splenda sugar substitute	4142.5	splenda sugar substitute	4142.5
Substitutos Adequados	5		4		5		5				
Receita		Quick Pound Cake		Simple Pudge Brownie		Wheat Germ Banana Muffin		Whole Wheat Corn Bread			
Ingrediente Alvo		flour		flour		all-purpose flour		wheat flour			
		Candidato	Escore	Candidato	Escore	Candidato	Escore	Candidato	Escore	Candidato	Escore
Sem Glúten	oil	3491.7		baking powder	3491.1	baking soda	3493.5	baking soda	3493.5	baking soda	3493.5
	baking powder	3489.4		oil	3489.1	zanthan gum	3489.5	zanthan gum	3489.5	zanthan gum	3489.5
	baking soda	3488.4		baking soda	3489.1	oil	3489.4	oil	3489.4	oil	3489.4
	zanthan gum	3487.4	3487.4	zanthan gum	3487.4	vanilla	3487.5	vanilla	3487.5	vanilla	3487.5
	oil	3487.0		rice flour	3486.4	rice flour	3486.4	rice flour	3487.1	rice flour	3487.1
	vanilla	3486.1		oil	3486.3	cinamon	3484.7	water	3485.2	water	3485.2
	rice flour	3485.0	3485.0	brown sugar	3483.6	water	3483.5	cinamon	3483.3	cinamon	3483.3
	water	3484.5		water	3483.1	potato starch	3483.0	potato starch	3483.0	potato starch	3483.0
	vanilla extract	3483.3		cinamon	3482.6	gluten-free flour	3482.3	cornstarch	3482.5	cornstarch	3482.5
	cinamon	3481.2		topioca flour	3481.3	brown sugar	3482.2	gluten-free flour	3482.3	gluten-free flour	3482.3
Substitutos Adequados	2		3		4		5				
Receita		Quick Pound Cake		Simple Pudge Brownie		Wheat Germ Banana Muffin		Whole Wheat Corn Bread			
Ingrediente Alvo		butter		butter		milk		milk			
		Candidato	Escore	Candidato	Escore	Candidato	Escore	Candidato	Escore	Candidato	Escore
Sem Lactose	water	4761.8		water	4761.1	baking soda	4763.7	baking soda	4763.7	baking soda	4763.7
	brown sugar	4759.5		cinamon	4759.7	water	4763.8	water	4763.8	water	4763.8
	cinamon	4758.0		brown sugar	4758.5	cinamon	4752.5	flour	4763.4	flour	4763.4
	oil	4754.5	4754.5	banana	4751.3	soymilk	4762.1	soymilk	4761.1	soymilk	4761.1
	vegetable oil	4753.3	4753.3	vegetable oil	4750.5	flour	4762.0	all-purpose flour	4760.3	all-purpose flour	4760.3
	canola oil	4749.9	4749.9	oil	4750.8	vanilla	4759.4	vanilla	4760.1	vanilla	4760.1
	cornstarch	4748.3		canola oil	4749.6	whole wheat flour	4758.3	soymilk	4760.0	soymilk	4760.0
	banana	4747.1		ground cinnamon	4747.6	canola oil	4756.6	vanilla extract	4757.2	vanilla extract	4757.2
	ground cinnamon	4746.9		soymilk	4744.2	brown sugar	4755.6	canola oil	4755.2	canola oil	4755.2
	vanilla	4746.8		cornstarch	4744.1	ground cinnamon	4753.0	brown sugar	4752.8	brown sugar	4752.8
Substitutos Adequados	3		4		2		2				

tornando as receitas geradas seguras nesse novo domínio. Como tendências para trabalhos futuros, pretende-se melhorar a análise da relevância dos ingredientes em cada um dos domínios de receitas estudados, de modo a evitar que ingredientes muito frequentes, porém pouco relevantes no domínio, acabem tendo grande impacto nas recomendações, gerando, desta forma, recomendações mais precisas e adequadas ao contexto das receitas investigadas. Também pretendemos expandir a análise proposta pelo acréscimo de mais receitas exemplo. Por fim, pretendemos analisar a elaboração de abordagens que façam a substituição de múltiplos ingredientes simultaneamente.

Agradecimentos

Os autores gostariam de agradecer à FACEPE, ao CNPq e à CAPES, agências de fomento brasileiras. Os autores gostariam de agradecer ao CPQD pelo apoio financeiro.

REFERENCES

- LAWRYNOWICZ, A., WRÓBLEWSKA, A., ADRIAN, W. T., KULCZYŃSKI, B., AND GRAMZA-MICHALOWSKA, A. Food recipe ingredient substitution ontology design pattern. *Sensors* 22 (3): 1095, 2022.
- MACQUEEN, J. ET AL. Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*. Vol. 1. California, USA, pp. 281–297, 1967.
- MAJUMDER, B. P., LI, S., NI, J., AND MCAULEY, J. Generating personalized recipes from historical user preferences. *arXiv preprint arXiv:1909.00105*, 2019.
- NIRMAL, I., CALDERA, A., AND BANDARA, R. D. Optimization framework for flavour and nutrition balanced recipe: A data driven approach. In *2018 5th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON)*. IEEE, pp. 1–9, 2018.
- Ooi, A., IBA, T., AND TAKANO, K. Ingredient substitute recommendation for allergy-safe cooking based on food context. In *2015 IEEE Pacific Rim Conference on Communications, Computers and Signal Processing (PACRIM)*. IEEE, pp. 444–449, 2015.
- PACIFICO, L. D. S., BRITTO, L. F. S., AND LUDERMIR, T. B. Ingredient substitute recommendation based on collaborative filtering and recipe context for automatic allergy-safe recipe generation. In *Proceedings of the Brazilian Symposium on Multimedia and the Web*. WebMedia '21. Association for Computing Machinery, pp. 97–104, 2021.
- PECUNE, F., CALLEBERT, L., AND MARSELLA, S. A socially-aware conversational recommender system for personalized recipe recommendations. In *Proceedings of the 8th International Conference on Human-Agent Interaction*. HAI '20. Association for Computing Machinery, pp. 78–86, 2020.
- SHINO, N., YAMANISHI, R., AND FUKUMOTO, J. Recommendation system for alternative-ingredients based on co-occurrence relation on recipe database and the ingredient category. In *2016 5th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI)*. IEEE, pp. 173–178, 2016.