Enhancing GISMo: Integrating Recipe Contexts into a Graph-Based Ingredient Substitution Module

Julie Loesch[1], David Schimmel[1], Funda Erdem[2], Deryanur Kalkavan[2], Zeki Bilgin[2] and Remzi Celebi[1]

[1]*Department of Advanced Computing Sciences, Maastricht University, Paul-Henri Spaaklaan 1, Maastricht, 6229 EN, Netherlands*
[2]*BEKO A.S., Istanbul, Turkey*

Abstract

To comply with dietary restrictions, individuals seek to replace ingredients in culinary recipes. The advanced model Graph-based Ingredient Substitution Module, called GISMo, recently was proposed to facilitate the learning of ingredient substitutions within recipes. We identify some potential improvements in GISMo's functionality for generating substitution recommendations based on recipe contexts by capturing various graph representations of recipe contexts, including ingredient-recipe relationships and cooking actions. Furthermore, we introduce a novel benchmark dataset comprising substitutions for specific recipes, validated by four domain experts. To assess the proposed improvements, we conduct experiments on both existing and new substitution datasets. Our findings demonstrate that the proposed changes in GISMo can augment the diversity of food substitution recommendations without compromising prediction quality.

Keywords

knowledge graph, graph neural network, ingredient substitution, recipe context

1. Introduction

Selecting ingredient substitutions in culinary recipes is a balancing act between culinary art and science. Many factors that influence these selections are food availability, health considerations (including allergies or nutritional needs), personal taste preferences, ethical concerns (such as animal welfare or environmental impact), or simply curiosity and experimentation. Furthermore, certain ingredient substitutions in recipes can promote the adoption of new dietary patterns or regimes. Identifying ideal substitutions for recipes can promote the acceptance of new recipes, which can change dietary behaviors and can also promote healthier eating.

This work aims to understand the relationship between the context of recipes and the suitability of ingredient substitutions. The recipe context can be either *external* (such as user ratings or known food substitutions [1]) or *internal* (such as cooking instructions [2] and nutritional content [1]). Recent studies using external recipe contexts employ statistical

SWAT4HCLS 2025: The 16th International Conference on Semantic Web Applications and Tools for Health Care and Life Sciences, February 24–27, 2025, Barcelona, Spain
julie.loesch@maastrichtuniversity.nl (J. Loesch); david.schimmel@student.maastrichtuniversity.nl (D. Schimmel); funda.erdem@beko.com (F. Erdem); deryanur.kalkavan@beko.com (D. Kalkavan); zeki.bilgin@beko.com (Z. Bilgin); remzi.celebi@maastrichtuniversity.nl (R. Celebi)
0000-0001-7769-4272 (R. Celebi)

methods, such as the co-occurrence of ingredients in recipe databases [3, 4, 1, 5]. Other studies capture latent representations from recipes by creating ingredient embeddings using language models [2, 6]. Furthermore, several works also explore graph-based approaches, integrating information about recipes and ingredients through ontologies [7, 8, 9, 10, 11, 12] or constructing directed acyclic graphs representing the structure of cooking instructions [12].

Our study investigates whether incorporating semantics of recipes enhances the performance of one of the state-of-the-art graph neural networks for ingredient substitutions, called Graph-based Ingredient Substitution Module (GISMo) [8]. This model was selected for its superior performance compared to other contemporary methods for identifying ingredient substitutions, such as FoodBert or Food2Vec ingredient embeddings [2].

Our primary contributions are summarized as follows:

- Creation of a novel ground truth dataset, ArcSubs, for ingredient substitutions within the context of specific recipes. Unlike previously published datasets, this new dataset is validated by domain experts.
- Evaluation of different input data representations in GISMo.
- Demonstrating increased diversity of recommendations for a source ingredient by encoding additional recipe information into the model's input graph representations.

2. Related Work

Related studies addressing the issue of ingredient substitutions can be classified into two primary categories. In the first category, the works concentrate on identifying effective substitutions irrespective of the recipe context. In the second category, studies aim to identify substitutions that are also suitable for the context of a specific recipe.

What constitutes a ”good” substitution often revolves around its acceptability to the consumer, though additional or alternative constraints can be imposed. For instance, substitutions may be constrained to offer a *healthier* alternative to the original ingredient [9, 10, 11]. In certain circumstances, substitutions are mandated to align with specific dietary preferences, such as transforming a non-vegan recipe into a vegan one, or altering the style or cuisine of the dish [13, 14].

When recipe information is represented as a knowledge graph, a common approach for retrieving substitutions involves knowledge graph embedding algorithms like TransE [15]. Studies including Loesch et al. [9, 10], Park et al. [7], Shirai and Kim [12] use KG embeddings and compare node embeddings to identify suitable substitutions. Alternatively, triple or link prediction tasks are employed to predict ingredient candidates that align with a recipe represented as a graph [16]. Ławrynowicz et al. [17] proposes ontology design patterns tailored for modeling ingredient substitutions in culinary recipes. To our knowledge, GISMo stands out as the only graph neural network explicitly designed to identify substitutions of ingredients for given recipes [8].

3. Data Collection

This section explains the generation of a new ground truth dataset for recipe sensitive ingredient substitutions. This dataset was collected through a survey conducted among four domain experts working at Arçelik[1]. We first describe the survey design, followed by the process of conducting the survey. Finally, we evaluate the results and present the ground truth dataset.

3.1. Survey Design and Process

The survey was designed to capture various dimensions of the ingredient substitution problem. The participants were provided with the recipe title, ingredient set with quantities, cooking instructions, and a suggested ingredient substitution pair. To generate ingredient substitution pairs for each recipe, we relied on Recipe1MSubs [18], a benchmark for ingredient substitution associated with Recipe1M [19] recipes. For each recipe in the survey, the following questions were asked:

- Question 1: Given the suggested substitution, does the suggested substitution fit the recipe? [Yes/No]
- Question 2: Given the suggested substitution, does this substitution imply a major change to the meal's taste? [Yes/No]
- Question 3: Given the suggested substitution, does this substitution imply a major change to the meal's nutritional profile? [Yes/No]
- Question 4: Given the suggested substitution, does this substitution require major modifications to the cooking process (i.e., instruction set)? [Yes/No]
- Question 5: Given the suggested substitution, does the substitution change the food category from the exchanged ingredient (i.e., swapping a meat for a vegetarian option or changing ingredients that do not cause the same cross-allergies)? [Yes/No]

3.2. Survey Evaluation

The total average number of responses per question, excluding Question 3, is approximately 3.89. The participants indicated that they were not comfortable answering Question 3 without additional information. To evaluate the survey responses, we computed the Fleiss-Kappa score, a statistical measure used to assess inter-rater reliability or agreement among multiple annotators. Question 1 achieved the highest clarity with a score of 0.73, indicating strong agreement among participants. In contrast, Question 2 exhibited poorer agreement among annotators, potentially due to the subjective nature of taste perception. Question 4 and Question 5 obtained a Fleiss-Kappa score of 0.58 and 0.67, respectively.

By computing the tetrachoric correlation among Questions 1, 2, 4, and 5, we noticed a weak negative correlation between the assessment of whether a substitution fits the recipe and whether it alters the taste of the dish. This suggests that while a substitution may fit the recipe, it often affects the taste profile of the recipe. The most pronounced negative correlation appears between Questions 1 and 4, indicating that substitutions deemed fitting are less likely to require significant modifications to the cooking process.

[1]https://www.arcelikglobal.com/en/company/about-us/overview/

3.3. Ingredient Substitution Benchmark Dataset

The benchmark includes all samples where a majority of annotators agree that the suggested substitution fits the recipe, regardless of any changes in taste. Additionally, participants were asked to choose their preferred substitute ingredient from three options or propose a more suitable alternative. Responses to this question were used to enrich the ground truth samples. Majority voting for a particular option indicates a consensus that the ingredient fits the recipe. The final benchmark dataset, ArcSubs, comprises a total of 656 samples, spanning 448 distinct recipes and 182 distinct substitution pairs.

4. Methodology

We propose to improve GISMo by addressing two key aspects: graph structure and diversity. Specifically, the input representation for GISMo, FlavorGraph, links ingredient nodes based on their normalized point-wise mutual information, calculated from their co-occurrences in the dataset. However, these connections do not necessarily imply good substitutions, as they can also represent complements [7]. Our preliminary experiments showed that the removal of 20% of these edges did not affect the model's prediction capability, suggesting that the current input structure does not contribute well to predict good substitutions. In this section, we introduce different input data representations for GISMo by incorporating additional contextual information.

4.1. Bipartite Graph of Ingredients and Recipes

To improve GISMo, we propose to remove all direct connections between ingredient nodes in FlavorGraph. Furthermore, we introduce a new type of node representing recipes. These recipe nodes are connected to ingredient nodes through undirected edges, indicating the inclusion of the ingredient in the recipe. This new structure forms a bipartite graph, where recipe nodes are linked to their respective ingredients, thereby capturing the context of recipes through their ingredient sets. This bipartite graph representation is expected to enhance the model's ability to learn more accurate and contextualized ingredient representations. Figure 1 illustrates this revised graph schema.

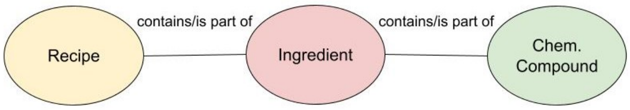

Figure 1: Schema for bipartite graph including recipe-ingredient context.

4.2. Arbitrary Ring Graph

To further assess the impact of modifications to the graph structure, we create an experimental graph by first removing all existing edges in FlavorGraph. We will then add arbitrary edges, forming a ring structure that sequentially connects all nodes in a circle. Additionally, we will introduce chords that connect every n-th node to every other n-th node, with the exception that every n-th non-ingredient node will only connect to every n-th ingredient node, leaving part of the circle without interconnected chords. This ring structure will help to evaluate the effect of arbitrary and structured connections on the model's performance.

4.3. Bipartite Graph of Ingredients, Recipes, and Action Nodes

Incorporating partially structured context of cooking instructions can provide additional valuable information for finding substitutions. To achieve this, we extracted verbs and nouns from cooking instructions given in natural language sentences, as in the work of [12]. In Recipe1M, instructions are provided in an ordered list of steps. Each step consists of at least one verb and zero or more nouns. To extract nouns and verbs from these steps, we used a pre-trained transformer model provided by spaCy [20]. We assume that the extracted verbs represent distinct cooking actions.

Each cooking action is represented by a node and is connected to a recipe node if the cooking action is part of the recipe. This graph is schematized in Figure 2. Intuitively, this graph does not represent the structure of the cooking instructions. Instead, it models the relationship between ingredients and cooking actions, indicating which cooking action is applied to which ingredient. The importance of the relationship between cooking actions and ingredients for finding substitutions has been emphasized by several studies, such as in the work of [21].

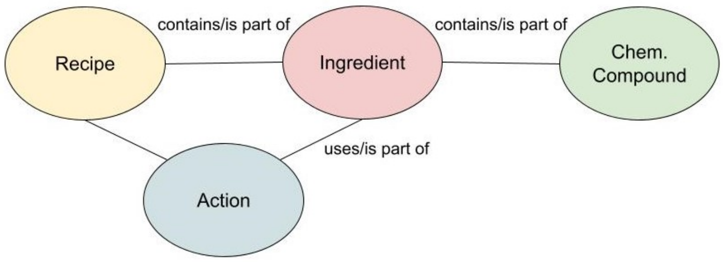

Figure 2: Schema with recipe context including cooking actions.

4.4. Flow Graph

In our final graph representation, we aim to generate a structure that reflects the detailed cooking process inspired by [12]. The primary challenge in this representation is managing the branching points, which occur when mutually independent sub-sequences result in the same intermediary food product. For example, components like cake, pie crust, and topping may be processed independently, even though each follow a specific sequence of instructions. Ideally, a graph representing the cooking process would represent these substructures separately, creating distinct branches that converge to the shared outcome.

The aim is to represent an approximation of the dependencies inherent in the cooking instructions, thereby forming a Directed Acyclic Graph (DAG) that illustrates the progression from raw ingredients to the final food product. This comprehensive structure aims to enhance GISMo's ability to learn contextualized ingredient representations by accurately modeling the sequence, dependencies of cooking actions and ingredient transformations, thereby improving substitution recommendations. The schema for the graph generated in this fashion is illustrated in Figure 3.

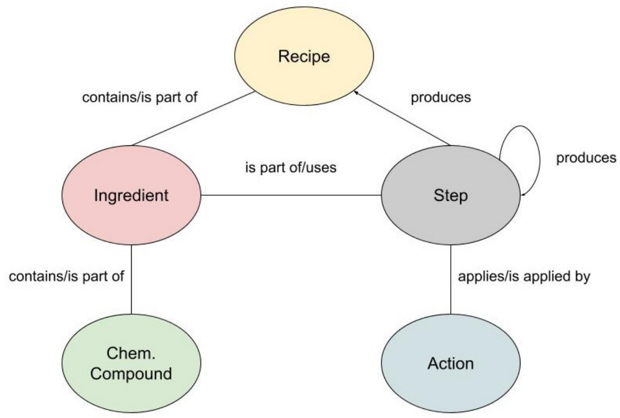

Figure 3: Schema for recipe context with intermediary step nodes.

5. Experiments

To assess various input data representations for GISMo, we experimented with enhanced graph representations to prioritize potential substitution targets for given recipes and substitution sources.

- Graph-based Ingredient Substitution Module (*GISMo*): GISMo [8] serves as the baseline model.
- Bipartite Graph of Ingredients and Recipes (*Bipartite*): This input graph is a bipartite graph comprising nodes for recipes and ingredients, connected if the ingredients are used in the recipe.
- Arbitrary Ring Graph (*Ring*): In this model, the input graph retains the same nodes as GISMo but substitutes the edges with arbitrary connections.
- Bipartite Graph of Ingredients, Recipes, and Action Nodes (*Bipartite + Actions*): This input graph is a bipartite graph involving nodes for recipes, ingredients, and cooking actions. Recipes connect with cooking actions and ingredients based on their usage in the recipe. Ingredients and cooking actions are linked if they are involved in the same cooking step.
- Flow Graph (*Flow*): This model utilizes a graph where recipes are represented as flow graph-like subgraphs, with intermediary nodes corresponding to the products of each action in the cooking instructions connected by directed edges.

- Minimized Flow Graph (*Flow**): This model employs a simplified version of *Flow*, focusing only on recipes present in the ArcSubs dataset and represented as flow graph-like subgraphs.

One potential issue with the baseline model is that the number of distinct recommended substitutions in the top ranks is often much lower than the distribution of ground truth substitutions for certain source ingredients. Therefore, models were also assessed using rank diversity $d^{I_s^R}(k)$, which counts the distinct ingredient recommendations at rank k across all samples. Here, I_s^R denotes the source ingredient of the sample recipe R.

Additionally, we count the number of substitution sources that appear in more than 10 recipes and have only one top predicted substitution target. For example, if the model consistently predicts "ground turkey" as the top recommended substitution for the source ingredient "ground beef," and "ground beef" is a substitution source in more than 10 recipes, then $d(1)^{|R|>10} = 1$ increases by 1.

Hence, experiments were conducted on two datasets, Recipe1MSubs and ArcSubs, evaluating models using Mean Reciprocal Rank (MRR), Hits@k, rank diversity and we count the number of substitution sources that appear in more than 10 recipes.

6. Results

Table 1 and Table 2 provide an overview of the measured Mean Reciprocal Rank (MRR), Hits@k, and rank 1 diversity for the evaluated models. Table 1 details the results for Recipe1MSubs, while Table 2 presents the results for ArcSubs. The tables include findings for the baseline model *GISMo*, bipartite graphs *Bipartite* and *Bipartite + Actions*, and the ring graph with chords *Ring*. Due to high memory requirements, experiments with the flow graph-like representation *Flow* were not feasible; thus, results are only available for the minimized flow graph *Flow**. It is important to note that *Flow** encodes recipes exclusively from the ArcSubs dataset, hence its results are reported solely in Table 2.

Metric	*GISMo*	*Bipartite*	*Ring*	*Bipartite + Actions*
MRR	31.4	31.4	31.6	31.6
Hits@1	20.4	20.3	20.6	20.6
Hits@3	35.6	35.9	36.2	36.1
Hits@10	54.1	54.0	54.4	54.7
$d(1)$	1.4	1.4	1.2	1.5
$d(1)^{\|R\|>10} = 1$	68	58	53	52

Table 1
Results for Recipe1MSubs.

The bipartite graph with action nodes, *Bipartite + Actions*, demonstrates superior performance among the models with a mean reciprocal rank of 31.6 for Recipe1MSubs and 90.0 for ArcSubs datasets. Additionally, it achieved the highest rank diversity $d(1)$ for Recipe1MSubs, indicating that incorporating recipe context with ingredient and action nodes improves *GISMo*.

Metric	*GISMo*	*Bipartite*	*Ring*	*Bipartite + Actions*	*Flow**
MRR	86.2	81.1	86.1	90.0	85.3
Hits@1	79.7	68.3	73.3	81.7	73.3
Hits@3	98.4	93.3	100.0	100.0	100.0
$d(1)$	1.1	1.0	1.1	1.0	1.5
$d(1)^{\vert R\vert>10} = 1$	0	0	0	0	52

Table 2
Results for ArcSubs.

From the tables, it is evident that the bipartite graph *Bipartite*, which includes recipe and ingredient nodes, performs notably worse than the baseline model *GISMo*, particularly on the ArcSubs dataset. The arbitrary ring graph *Ring* and the minimized flow graph *Flow** show results similar to *GISMo* across both datasets.

7. Discussion

The results indicate that various input data representations can improve GISMo's performance. Among the proposed changes, the bipartite graph incorporating ingredient, recipe, and action nodes (*Bipartite + Actions*) showed particularly promising results and an increased diversity of ingredient recommendations.

Our analysis compared the outputs of *Bipartite + Actions* and GISMo, revealing a notable difference: while the baseline model consistently recommended ”butter” as the top substitution for ”olive oil,” the enhanced bipartite graph suggested four distinct alternatives—”butter,” ”apple sauce,” ”canola oil,” and ”sesame oil.” All of these recommendations seem plausible, depending on the recipe's context. For instance, sesame oil could be an excellent substitute when a nutty aroma is desired, whereas canola oil or butter might be more appropriate for recipes requiring a neutral flavor. The increased diversity in substitutions also supports specific goals, such as finding vegan alternatives. In such cases, butter would not be suitable, and a more varied selection of substitutions would be advantageous.

However, it's important to note that these enhancements come with increased model space requirements and longer training times. The additional computational cost should be carefully balanced against the observed improvements if these changes were to be implemented in a production environment.

8. Conclusion

This work investigates whether incorporating semantic information about recipe components enhances the performance of the state-of-the-art Graph-based Ingredient Substitution Module (GISMo) [8] in the identification of suitable ingredient substitutions within food recipes. We improved the model by integrating the pertinent recipe context into the model's input. Various graph structures were proposed and evaluated using two datasets: Recipe1MSubs, a standard for ingredient substitution associated with Recipe1M [19] recipes, and ArcSubs, a novel dataset where ingredient substitutions were evaluated by four domain experts within specific recipe contexts.

While traditional ranking-based metrics may not show substantial improvements compared to the baseline model GISMo, we observed that incorporating additional recipe context can increase the diversity of substitution recommendations. Among the proposed changes, the bipartite graph incorporating the ingredient, recipe and action nodes (*Bipartite + Actions*) showed promising results. The increased diversity, alongside maintained prediction quality measured by metrics such as mean reciprocal rank and Hits@k, suggests that the model can offer more tailored ingredient substitution recommendations aligned with the specifics of each recipe's context.

Acknowledgment

Thanks to Arçelik https://www.arcelik.com.tr, for supporting us in the survey data collection and for funding this work.

Online Resources

The adaptations to the investigated model can be found here: https://github.com/DavidSchimmel/gismo/tree/um_mt_extensions. The processing pipelines for generating the different graph representations can be found here: https://github.com/DavidSchimmel/structured_recipe1m. The remaining utility scripts, notebooks and other scripts related to this project are provided here: https://github.com/DavidSchimmel/statistical_ingredient_substitutions.

Declaration on Generative AI

We have used ChatGPT to address the grammatical errors and rephrase the sentences.

References

[1] S. Akkoyunlu, C. Manfredotti, A. Cornuéjols, N. Darcel, F. Delaere, Investigating substitutability of food items in consumption data, in: Second International Workshop on Health Recommender Systems co-located with ACM RecSys, volume 5, 2017.

[2] C. Pellegrini, E. Özsoy, M. Wintergerst, G. Groh, Exploiting food embeddings for ingredient substitution., in: HEALTHINF, 2021, pp. 67–77.

[3] R. Yamanishi, N. Shino, Y. Nishihara, J. Fukumoto, A. Kaizaki, Alternative-ingredient recommendation based on co-occurrence relation on recipe database, Procedia Computer Science 60 (2015) 986–993.

[4] N. Shino, R. Yamanishi, J. Fukumoto, 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, 2016, pp. 173–178.

[5] Y. Song, X. Yang, C. Xu, Self-supervised calorie-aware heterogeneous graph networks for food recommendation, ACM Transactions on Multimedia Computing, Communications and Applications 19 (2023) 1–23.

[6] A. Morales-Garzón, J. Gómez-Romero, M. J. Martin-Bautista, A word embedding-based method for unsupervised adaptation of cooking recipes, IEEE Access 9 (2021) 27389–27404.

[7] D. Park, K. Kim, S. Kim, M. Spranger, J. Kang, Flavorgraph: a large-scale food-chemical graph for generating food representations and recommending food pairings, Scientific reports 11 (2021) 931.

[8] B. Fatemi, Q. Duval, R. Girdhar, M. Drozdzal, A. Romero-Soriano, Learning to substitute ingredients in recipes, arXiv preprint arXiv:2302.07960 (2023).

[9] J. Loesch, L. Meeckers, I. van Lier, A. de Boer, M. Dumontier, R. Celebi, Automated identification of food substitutions using knowledge graph embeddings., in: SWAT4HCLS, 2022, pp. 19–28.

[10] J. Loesch, I. van Lier, A. de Boer, J. Scholtes, M. Dumontier, R. Celebi, Automated identification of healthier food substitutions through a combination of graph neural networks and nutri-scores, Journal of Food Composition and Analysis 125 (2024) 105829. doi:https://doi.org/10.1016/j.jfca.2023.105829.

[11] S. S. Shirai, O. Seneviratne, M. E. Gordon, C.-H. Chen, D. L. McGuinness, Identifying ingredient substitutions using a knowledge graph of food, Frontiers in Artificial Intelligence 3 (2021) 621766.

[12] S. S. Shirai, H. Kim, Eat-pim: Substituting entities in procedural instructions using flow graphs and embeddings, in: International Semantic Web Conference, Springer, 2022, pp. 161–178.

[13] M. Kazama, M. Sugimoto, C. Hosokawa, K. Matsushima, L. R. Varshney, Y. Ishikawa, A neural network system for transformation of regional cuisine style, Frontiers in ICT 5 (2018) 14.

[14] Y. Pan, Q. Xu, Y. Li, Food recipe alternation and generation with natural language processing techniques, in: 2020 IEEE 36th International Conference on Data Engineering Workshops (ICDEW), IEEE, 2020, pp. 94–97.

[15] A. Bordes, N. Usunier, A. Garcia-Durán, J. Weston, O. Yakhnenko, Translating embeddings

for modeling multi-relational data, in: Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2, NIPS'13, Curran Associates Inc., Red Hook, NY, USA, 2013, p. 2787–2795.

[16] D. Li, M. J. Zaki, Food knowledge representation learning with adversarial substitution, in: Proceedings of the 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing, 2022, pp. 653–664.

[17] A. Ławrynowicz, A. Wróblewska, W. T. Adrian, B. Kulczyński, A. Gramza-Michałowska, Food recipe ingredient substitution ontology design pattern, Sensors 22 (2022) 1095.

[18] B. Fatemi, Q. Duval, R. Girdhar, M. Drozdzal, A. Romero-Soriano, Learning to substitute ingredients in recipes, 2023. arXiv:2302.07960.

[19] J. Marín, A. Biswas, F. Ofli, N. Hynes, A. Salvador, Y. Aytar, I. Weber, A. Torralba, Recipe1m: A dataset for learning cross-modal embeddings for cooking recipes and food images, CoRR abs/1810.06553 (2018). URL: http://arxiv.org/abs/1810.06553. arXiv:1810.06553.

[20] M. Honnibal, I. Montani, S. Van Landeghem, A. Boyd, et al., spacy: Industrial-strength natural language processing in python (2020).

[21] Y. Shidochi, T. Takahashi, I. Ide, H. Murase, Finding replaceable materials in cooking recipe texts considering characteristic cooking actions, in: Proceedings of the ACM multimedia 2009 workshop on Multimedia for cooking and eating activities, 2009, pp. 9–14.