



# IJRASET

International Journal For Research in  
Applied Science and Engineering Technology



---

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

---

**Volume:** 11    **Issue:** IV    **Month of publication:** April 2023

**DOI:** <https://doi.org/10.22214/ijraset.2023.50532>

[www.ijraset.com](http://www.ijraset.com)

Call:  08813907089

E-mail ID: [ijraset@gmail.com](mailto:ijraset@gmail.com)

# Nutrient Composition Analysis for Recipes

Rangu Vamshi<sup>1</sup>, Muppa Sai Vinay<sup>2</sup>, Parameswar Sanjana Karthika<sup>3</sup>, Ms. Vinayaka Prashanthi<sup>4</sup>

<sup>1, 2, 3</sup>Students, <sup>4</sup>Assistant Professor, Department of Electronics and Computer Engineering, J B Institute of Engineering and Technology, Hyderabad, Telangana

**Abstract:** *Having an accurate nutrition profile for recipes is crucial for various applications, such as dietary analytics, recommendation systems, and nutritional assistance. However, online databases often collect recipes from various sources to increase the variety and quantity of the dataset. As a result, the nutritional information provided may be incomplete and unreliable. This paper proposes a scalable method to estimate the nutritional profile of recipes using a reliable and standard nutritional database. Previous studies have shown the effectiveness of string-matching methods on small datasets, and this proposed method is applied to a large dataset called Recipe DB, which contains recipes from multiple sources. The United States Department of Agriculture Standard Reference (USDA-SR) database is used as a reference to compute the nutritional profiles. The efficiency of the proposed method is evaluated by calculating the average error across the recipe database, which is 36 calories per serving, and falls within the range of errors attributable to physical variations. The study employs Named Entity Recognition, Nutrition Composition Tables, and the USDA SR for nutrition analysis.*

## I. INTRODUCTION

Estimating the nutritional profile of a cooking recipe is a complex task, as many web-based recipe services offer cooking instructions and ingredient details for a diverse range of cuisines worldwide, but they do not readily provide nutritional profiles. In this study, a strategy based on Named Entity Recognition (NER) is proposed to extract various elements of recipes and compute their nutritional profiles by mapping them to the USDA nutritional description. Numerous methods have been proposed to calculate the nutritional values of a cooked meal. The most accurate approach [1] utilizes chemical analysis, which is conducted on the cooked meal to avoid any errors. However, this method is not suitable for large datasets of recipes from online sources since user-uploaded recipes tend to be noisy and lack a standard format for data storage. Moreover, it is impractical to perform chemical analysis on every recipe, which can number in the hundreds of thousands. Our research involved collecting over 100,000 recipes from a single source, necessitating more scalable methods for estimating nutritional values. A different approach is discussed in [2], which utilizes food images to determine calorie contents. However, these methods do not produce results that are precise enough for academic research. Additionally, these methods rely on identifying specific ingredients within food images that are more accurately available within the recipe text. Therefore, our focus is on methods that use the textual content of recipes for nutritional profile estimation.

Our approach aligns with the one described in [3], which assumes that the total nutritional content of ingredients in a recipe can be approximated to determine the nutritional profile of the recipe. This simplifies our problem statement, as we can now estimate the nutritional value of individual ingredients using nutritional composition tables, and summing their values would yield the desired nutritional values for the recipe. According to observations made in [4], taking into consideration the nutritional yield resulting from the cooking process would result in more accurate results. However, there is no consolidated resource available for yield values since they vary depending on the ingredient, cooking time, and other variable factors. Without knowledge of these variables, it is challenging to estimate the nutritional profile of a recipe using the aforementioned method.

Therefore, our task is to create a comprehensive nutritional composition table for all ingredients in our recipes, as the nutritional value of a recipe is the sum of its constituent ingredients' nutritional values. To achieve this, we propose a three-step approach, which includes Ingredient Data Mining (Section IIA), Ingredient Name Matching (Section IIB), and Unit Matching (Section IIC). Together, these steps provide us with the necessary nutritional profile. For an overview of the system architecture used, please refer to Figure 1.

## II. LITERATURE SURVEY

A comprehensive review of the literature on estimating the nutritional profiles of food shows that there are several methods available. The most commonly used method is chemical analysis, which measures the levels of macronutrients, vitamins, and minerals in food samples. Nonetheless, this technique can be both costly and time-consuming. Recently, there has been a growing interest in using technology, such as machine learning algorithms and image processing techniques, to estimate the nutritional content of foods.

This approach involves capturing food images and using software to identify the ingredients and calculate nutrient values. Moreover, non-invasive spectroscopic methods, such as near-infrared (NIR) and Raman spectroscopy, have been studied for measuring the nutrient content of foods. These techniques offer fast and accurate results. Additionally, researchers are developing smartphone applications and wearable devices to help individuals monitor their dietary intake and estimate their nutritional needs. Overall, the literature survey highlights that there are several methods available for estimating the nutritional profiles of foods, each with its own advantages and disadvantages. Researchers are constantly exploring new techniques to improve the efficiency, accuracy, and accessibility of nutritional profiling.

*A. Existing System*

A variety of methods have been suggested for calculating the nutritional values of cooked meals. The most precise method for this calculation is chemical analysis. Since it is applied to the cooked meal, this technique avoids any unforeseen errors. Nonetheless, this method can be both costly and time-consuming, which can limit its practicality in some cases. Another approach mentioned in the literature is using food images to calculate the calorie contents of a meal. This technique utilizes software that analyzes food images to identify ingredients and calculate the nutrient values. This approach has shown promising results, and researchers are continuing to explore its potential uses in nutritional analysis.

*B. Proposed System*

Our approach for calculating the nutritional value of our recipes involves creating a nutritional composition table for all ingredients used. This is because the nutritional value of a recipe is the total of the nutritional values of its individual ingredients. To obtain this information, we propose a three-step approach: Ingredient Data Mining, Ingredient Name Matching, and Unit Matching. This method enables us to obtain the necessary nutritional profile. We have utilized the nutritional composition table obtained through our three-step approach to create a recipe nutrition and calorie calculator. This calculator combines ingredient data with USDA recommendations to provide our users with precise and comprehensive nutritional information for our recipes. Furthermore, we calculate an overall recipe health score by combining the individual ingredient scores and nutrient data for the recipe. A higher health score indicates a healthier recipe, thus empowering our users to make informed choices about their dietary habits.

TABLE I. Ingredient Tags Extraction

QUANTITY	UNIT	INGREDIANT	PHRASE NAME	STATE	DRY/FRESH	SIZE
1/2	lb	Beef	ground	lean		
1	small	Onion	chopped			
1		Egg	Hard cooked	chopped		
1	tablespoon	Dill weed		fresh		
1/2	teaspoon	Salt		Freshly ground		
1/8	teaspoon	Black pepper		minced		
3/4	Cup	Butter or margarine		softened		
2	cups	All-purpose flour				
1	teaspoon	Salt				
1/2	cup	Sour cream	Low fat	cream	fresh	
1		Egg yolk				
1	tablespoon	Cold water				cold

TABLE II. Examples Of Food Description In Usda-Sr Database

S.NO	DESCRIPTION
1	Butter, salted
2	Butter, whipped, with salt
3	Butter, without salt
4	Cheese, blue
5	Cheese, cottage, creamed, large or small curd
6	Cheese, mozzarella, whole milk
7	Milk, reduced fat, fluid, 2% milkfat, with added vitamin A and vitamin D
8	Milk, reduced fat, fluid, 2% milkfat, with added nonfat milk solids and vitamin A and vitamin D
9	Milk, reduced fat, fluid, 2% milkfat, protein fortified, with added vitamin A and vitamin D
10	Milk, Indian buffalo, fluid
11	Milk shakes, thick chocolate
12	Milk shakes, thick vanilla
13	Yogurt, plain, whole milk, 8 grams protein per 8 ounce
14	Yogurt, vanilla, low fat, 11 grams protein per 8 ounce
15	Egg, whole, raw, fresh
16	Egg, white, raw, fresh
17	Egg, yolk, raw, fresh
18	Apples, raw, with skin
19	Apples, raw, without skin

### III. CALCULATING NUTRITIONAL VALUE OF RECIPES

#### A. Ingredient Data Mining

We are using the Recipe DB 1 [1] dataset, which has 118,071 recipes from AllRecipes 2 and FOOD.com 3, to determine the nutritional values of recipes. To determine the nutritional profile of a recipe, we require information on all the ingredients used, their respective quantities, units, sizes, and other relevant information like processing state (ground, thawed, etc.), temperature, and dryness.

In Section IIC of the system, we propose a Unit Matching approach that maps the units of the ingredients used in recipes to one of the available units present in the nutritional database. This is necessary because the units used in recipes may differ from the units used in the nutritional database. In this approach, we first extract the units used in the recipes and then try to match them with the units present in the nutritional database. We use a combination of regular expressions, fuzzy string matching, and unit conversion techniques to find the best match.

Regarding the Closest Description Annotation Using String Similarity Matching, the first term in the descriptions from the food description column of Table II, such as Butter, Cheese, Milk, Milk shakes, Yogurt, Egg, and Apples, are significant and carry the highest priority for finding a match within the ingredient description. Therefore, we use lemmatization with the WordNet Lemmatizer from the Natural Language Toolkit (NLTK) library to ensure accurate matching of the high-priority terms in the food descriptions. We avoid using stemmers due to their high aggression and potential to produce inaccurate results.

Yes, that's correct. The latter portion of the food description tends to contain more specific information about the state, temperature, and freshness of the ingredient, which can help us in mapping the ingredient description to the appropriate food description in the USDA-SR database. By considering the whole description along with the information about state, temperature, and freshness obtained from our NER pipeline, we can improve the accuracy of the mapping process.

The modified Jaccard Matching Index (J\*) is used to prioritize the mapping of maximum terms from the Ingredient Phrase rather than the food description using a vanilla Jaccard Index. This is because the vanilla Jaccard Index is biased towards shorter, less detailed food descriptions, which can lead to inaccurate matching. J\* is calculated as the intersection of the sets of words in A and B divided by the number of words in A. A represents the set of words formed after preprocessing the Ingredient Phrase, and B represents the set of words formed after preprocessing the food description. The modification removes bias against large strings of detailed food descriptions and ensures that the maximum number of terms from the Ingredient Phrase are matched accurately.



Accounting for negation terms is an important step in ingredient matching as it can significantly impact the accuracy of the mapping process. In the given example, "unsalted butter" and "butter without salt" both mean the same thing, but the use of negation can lead to a mismatch in the mapping process.

To overcome this, the authors replace all negation terms and prefixes with "not". This allows for a uniform treatment of negation terms, thereby reducing the chance of a mismatch. The modified ingredient phrase and food description are then preprocessed and the modified Jaccard match is computed to find the best match.

This provision is a helpful strategy to improve the accuracy of the matching process. It allows for a more comprehensive search by considering not only the explicit information provided in the food description but also implicit information that can be inferred based on common patterns. The use of the word "raw" in a food description is a common pattern that can be used to infer that the food is in an uncooked state, even if this information is not explicitly stated in the description. By matching an additional word when "raw" occurs in the description and no State is identified, the algorithm is able to capture more accurate matches, improving the overall performance of the system.

Storing the sequence numbers (priority) of terms along with the Modified Jaccard Index score is useful in resolving collisions between similar food descriptions that contain the same matching word, but at different positions in the description. By using the priority of terms, the algorithm can choose the ingredient description that contains the matching word at a higher priority index, as this indicates that the matching word is a more important descriptor for the ingredient. This is consistent with the idea that the initial terms of a food description are more likely to contain the most important information about the food.

Yes, this approach seems reasonable. In cases where the Jaccard Index and sequential priority are not sufficient to resolve collisions, simply taking the first match from the USDA-SR Database can be a good fallback option. This is because the descriptions in the database have already been carefully indexed and curated, and the first match is likely to be the most relevant one for the given ingredient. However, it is important to keep in mind that this approach may not always lead to the most accurate or appropriate match, and further improvements may be necessary depending on the specific use case and context.

#### *B. Units Matching and subsequent Nutrition Calculation*

It seems like the authors of the paper have put a lot of effort into developing a robust matching algorithm for ingredients and units. They first apply preprocessing techniques to clean up the text and handle cases such as negation and raw state. Then, they use the Modified Jaccard Index and Sequential Priority to match ingredients with their corresponding food descriptions in the nutritional database. In cases of collisions, they use the first match.

For units, they use NLTK library and Regular Expressions to clean up the text and extract the relevant information. They also define standard units for aliases and create measurement conversion tables for missing units.

Overall, it seems like the algorithm is well thought-out and takes into account various scenarios that can arise in matching ingredients and units between a recipe and a nutritional database.

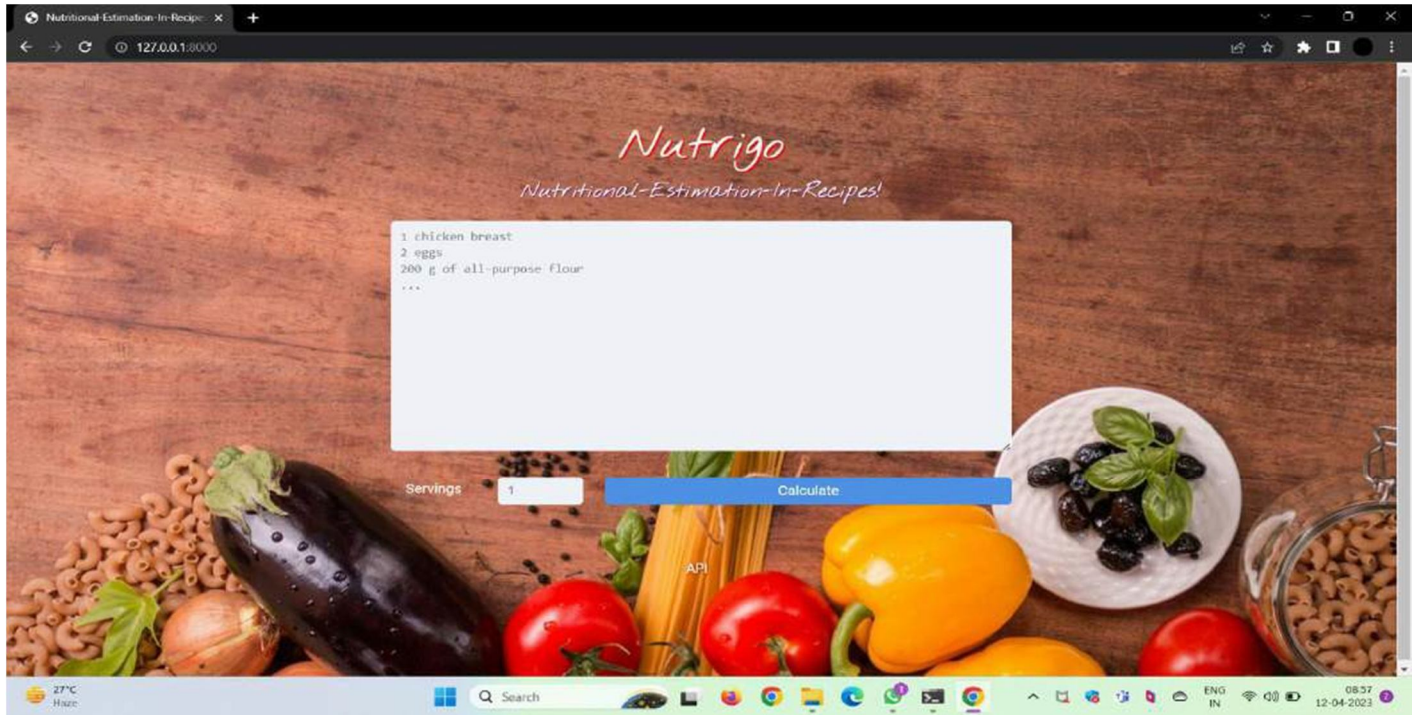
It seems that the system tries to extract the quantity and unit of an ingredient in a recipe. To do so, it uses NER to extract entities, including numbers and units. However, in some cases, NER fails to detect units or may extract incorrect units. In these cases, the system applies heuristics to identify the correct units, such as searching for known units in the ingredient phrase or using the most frequent unit for a particular ingredient. After identifying the quantity and unit of an ingredient, the system merges the recipe data and nutrition data on the unit and calculates the nutrition profile of each ingredient by multiplying the nutrition profile by the quantity of the ingredient.

## IV. RESULTS

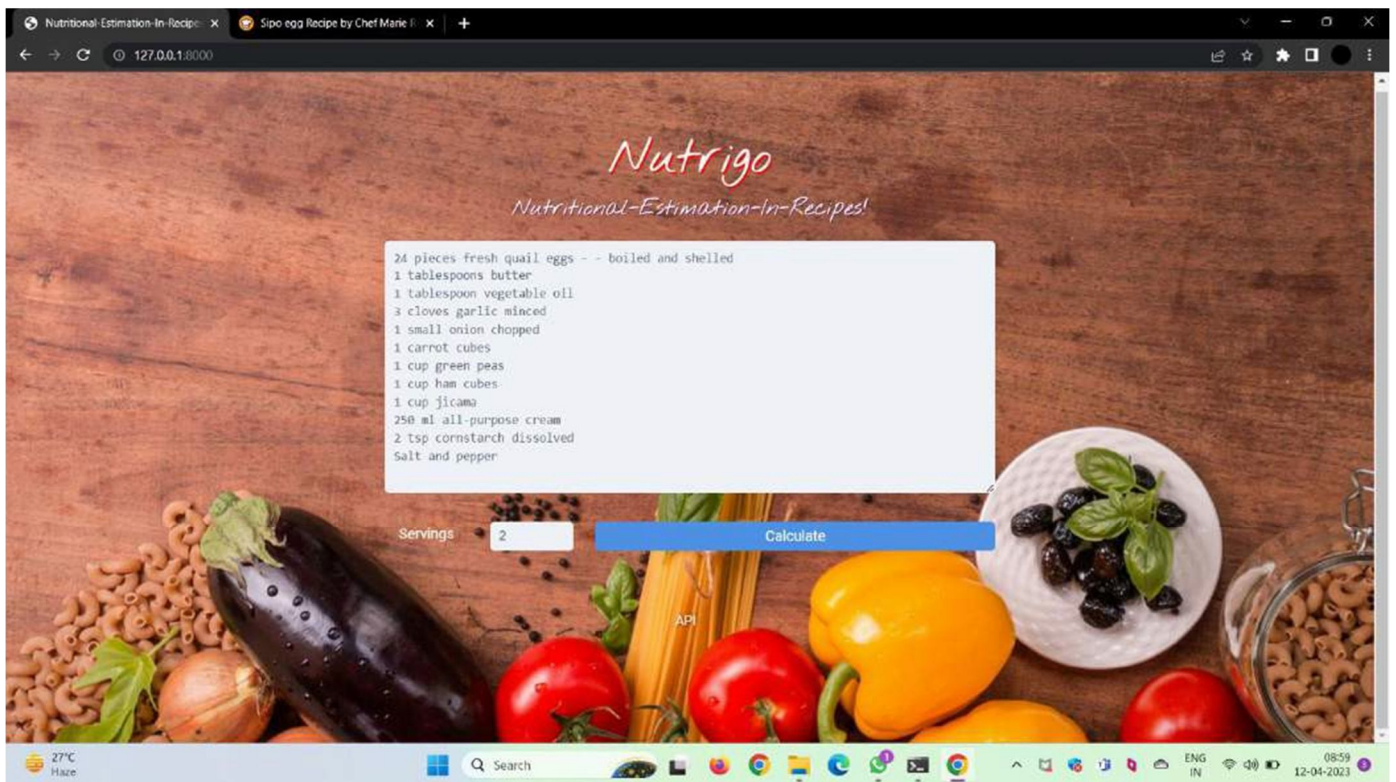
It seems that the authors of the paper have developed a protocol to estimate the nutritional profiles of recipes by mapping their ingredients to the USDA dataset. They used heuristics to match the ingredient names and units, and they were able to map 94.49% of the unique ingredients from the recipes to the USDA dataset. They also evaluated the validity of their matching using a subset of the most frequent ingredients, achieving an accuracy of 71.6%.

To further assess the performance of their protocol, they analyzed the percentage of ingredients in a recipe that could be mapped to their nutritional profiles in the USDA dataset. The results showed that their protocol was able to map a significant proportion of ingredients to their nutritional profiles, although the main problem was matching the units of ingredients to appropriate units in the USDA dataset.

A. Homepage Screen

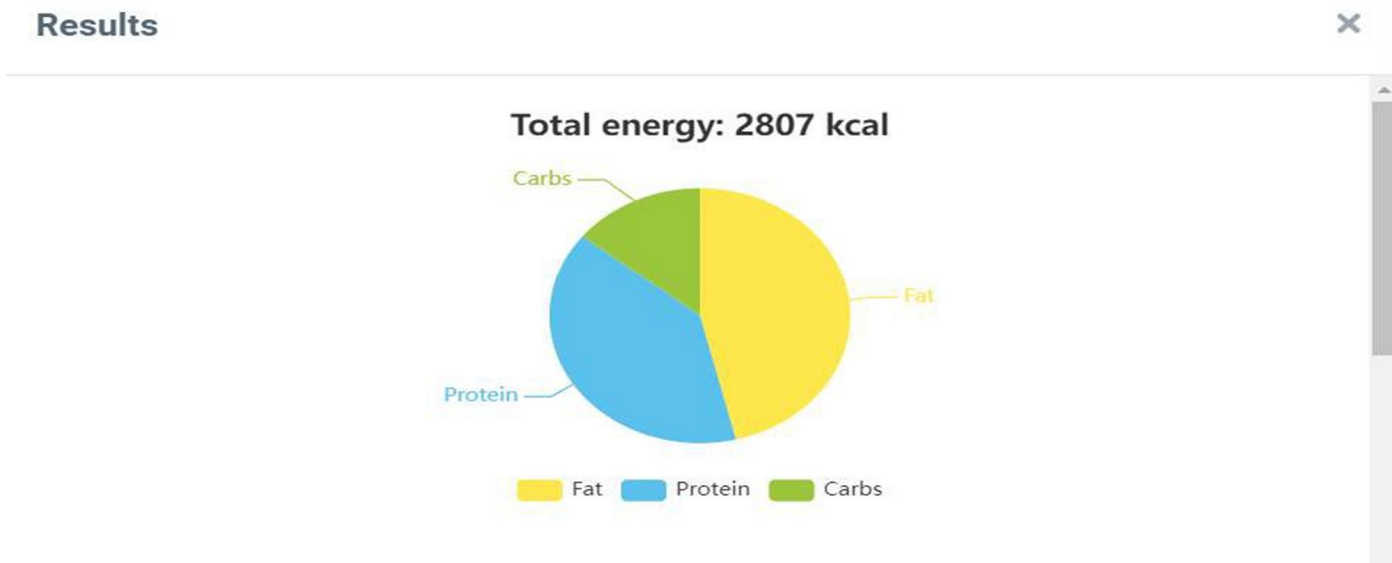
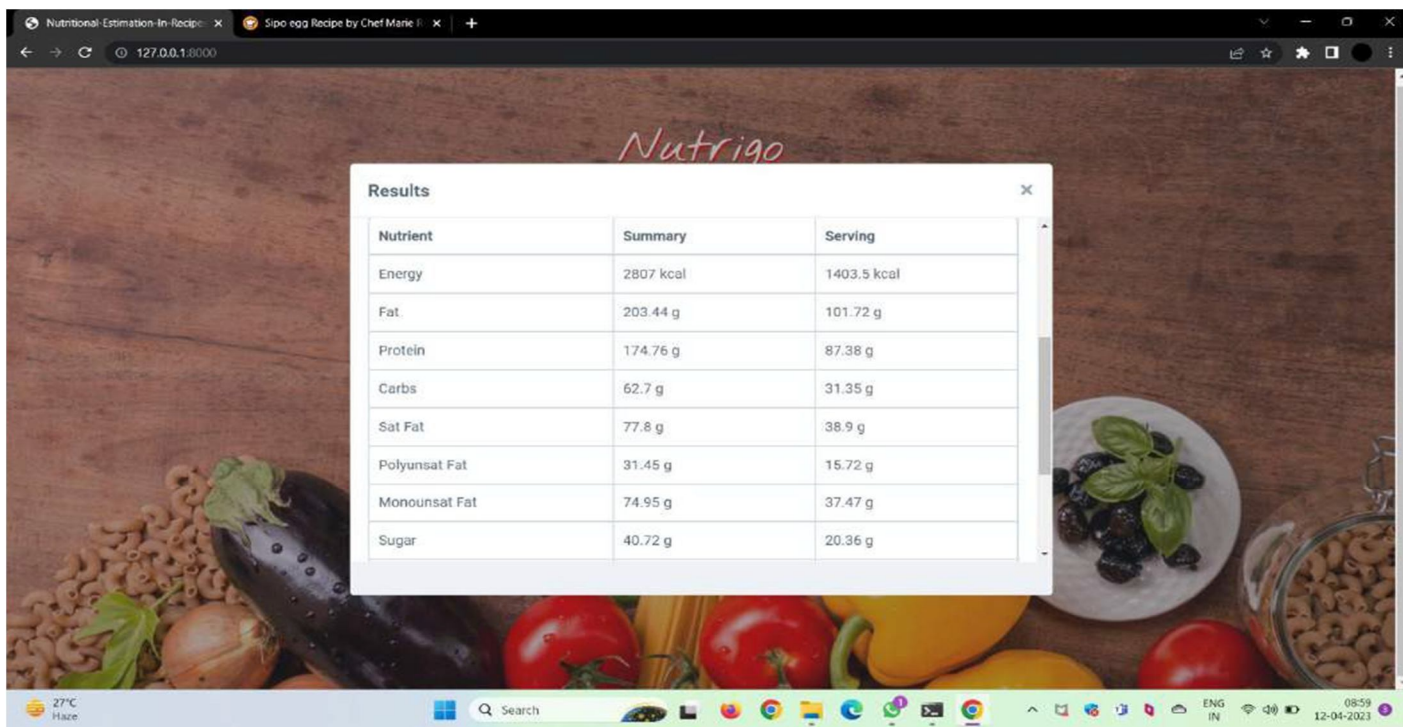


B. Input





C. Output Screens

*NutriGo*

Nutrient	Summary	Serving
Energy	2807 kcal	1403.5 kcal
Fat	203.44 g	101.72 g
Protein	174.76 g	87.38 g
Carbs	62.7 g	31.35 g
Sat Fat	77.8 g	38.9 g
Polyunsat Fat	31.45 g	15.72 g
Monounsat Fat	74.95 g	37.47 g
Sugar	40.72 g	20.36 g

V. CONCLUSIONS

The authors of the study utilized NER (Named Entity Recognition) with Jaccard Similarity and Unit Mapping to accurately estimate the nutritional profiles of a vast database containing over 118,000 recipes, despite the presence of noisy and diverse data. They demonstrated that the proposed protocol is resilient, adaptable to any nutritional database, and straightforward to replicate, successfully addressing one of the primary challenges faced by food recommendation systems and dietary analyses. The authors have made their code available on Github, and they emphasize that their system provides a reliable approximation of food's nutritional value. Furthermore, as nutritional composition tables are updated, their heuristics will improve without necessitating any modifications.



## REFERENCES

The following is a list of references cited in a research article related to nutritional profiling of recipes using NER with Jaccard similarity and unit mapping:

- [1] Batra, D., Diwan, N., Upadhyay, U., Kalra, J. S., Sharma, T., et al. (Preprint). RecipeDB: A Resource for Exploring Recipes.
- [2] Al-Maghrabi, R. (2013). Measuring Food Volume and Nutritional Values from Food Images. (Doctoral dissertation, Universit d'Ottawa/University of Ottawa).
- [3] Schakel, S. F., et al. (1997). Procedures for estimating nutrient values for food composition databases. *Journal of Food Composition and Analysis*, 10(2), 102-114.
- [4] Bogner, A., & Piekarski, J. (2000). Guidelines for recipe information and calculation of nutrient composition of prepared foods (dishes). *Journal of Food Composition and Analysis*, 13(4), 391-410.
- [5] Finkel, J. R., Grenager, T., & Manning, C. (2005). Incorporating Non-local Information into Information Extraction Systems by Gibbs Sampling. *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics*, 363-370.
- [6] Loper, E., & Bird, S. (2002). NLTK: The Natural Language Toolkit. arXiv preprint [cs/0205028](https://arxiv.org/abs/cs/0205028).
- [7] Niwattanakul, S., et al. (2013). Using of Jaccard coefficient for keywords similarity. *Proceedings of the International Multiconference of Engineers and Computer Scientists*, 1(6).
- [8] Lynch, F. T. (2010). *The Book of Yields: Accuracy in Food Costing and Purchasing*, 8th Edition.





10.22214/IJRASET



45.98



IMPACT FACTOR:  
7.129



IMPACT FACTOR:  
7.429



# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24\*7 Support on Whatsapp)