A Knowledge Graph for Semantic–Driven Healthiness Evaluation of Online Recipes[1]

Charalampos Chelmis [a,*] and Bedirhan Gergin [b]
[a] *Computer Science, State University of New York at Albany, NY, USA*
E-mail: cchelmis@albany.edu
[b] *Computer Science, State University of New York at Albany, NY, USA*
E-mail: bgergin@albany.edu

Abstract. The proliferation of recipes on the Web presents an opportunity for developing AI methods to promote healthy nutrition of people using the Internet as a source of food inspiration. Recent research endeavors have resulted in the development of ontologies related to food, and algorithmic solutions for ingredient substitution. However, there is a lack of a resource oriented towards promoting research in semantic-based algorithmic meal plan recommendation and/or individual ingredient substitution that explicitly incorporates healthiness into the recommendation process. To address this gap, we present a knowledge graph comprising a large collection of recipes sourced from Allrecipes.com, their ingredients and corresponding nutritional information, social interactions metadata, and healthiness information calculated based on two international nutritional standards. We describe the construction process of our knowledge graph, and show its utility in quantitatively evaluating the healthiness of online recipes.

Keywords: Data resourse, Web and public health, Ontology

1. Introduction

With more and more people turning to the Web for inspiration on what to cook [1], nutritional facts and well–being considerations are becoming equally important to factors such as preparation time and ease of execution in the selection process of recipes [2], particularly for those that are conscious about their overall health or body image [3, 4]. In turn, food recommendation systems have recently started incorporating health and dietary restrictions, and ingredient preferences in their recommendation process (e.g., [5–7]). At the same time, national health organizations around the world promote healthy, balanced diets for people to maintain and/or improve their overall health [8] due to the fact that causal connections between certain dietary choices and diseases, including cardiovascular diseases [9], stroke [10], cancer [11], and mental health [12], have been established.

Nutrition labels on prepackaged food, such as the one illustrated in Fig. 1a, strive to make the nutritional value of a product (e.g., the amount of fiber, fat, sugar and salt that one would consume by eating a certain food item) available to consumers. Nutritional information is even provided for recipes found on the Web (see Fig. 2). Unfortunately however, in both online recipes and food labels, daily recommended values are reported solely for a female between $19 - 30$ years of age with a daily intake of $2,000$ kCal. Without any guidance on how to properly interpret this information, or how to convert to one's own gender, age, and daily calorific intake, nutrition labels often confuse, rather than inform, customers about the nutritional value of a meal.

[1]Supported by the National Science Foundation.
[*]Corresponding author. E-mail: cchelmis@albany.edu.

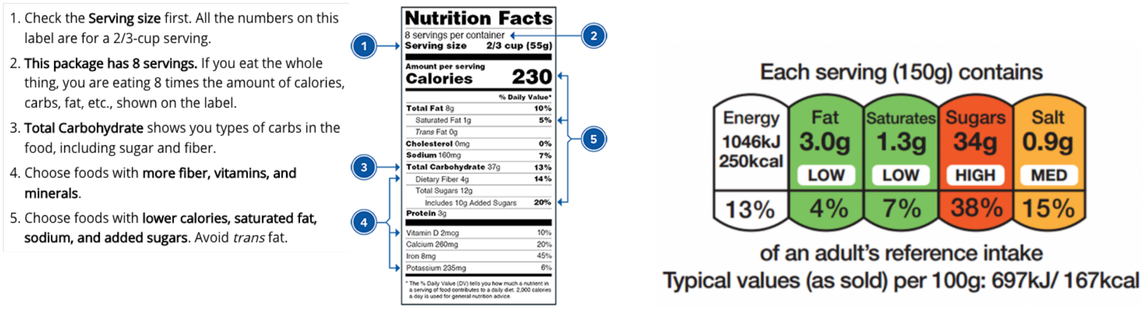


Fig. 1. (a) Sample nutrition facts label obtained from https://www.cdc.gov/diabetes/managing/eat-well/food-labels.html. Food labels such as this provide the amount of calories, carbs, fat, fiber, protein, and vitamins per serving of food, with the % daily value of each nutrient being calculated for a female between $19 - 30$ on a $2,000$ calories a day diet. (b) Sample traffic light label. Source: https://www.food.gov.uk/safety-hygiene/check-the-label.

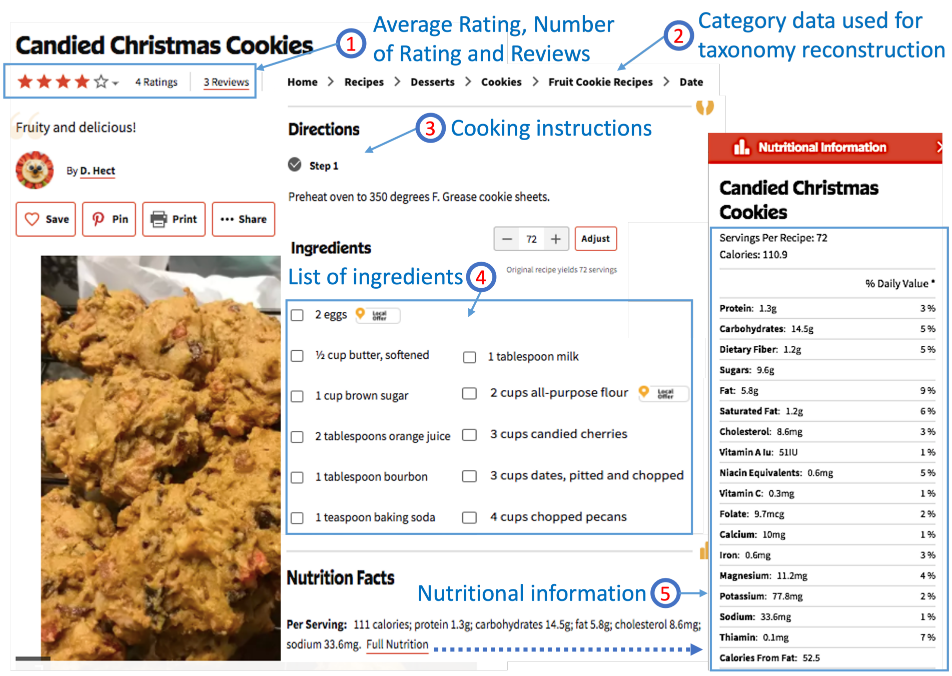


Fig. 2. Sample recipe from www.allrecipes.com. Apart from the list of ingredients and instructions on how to execute the recipe, nutritional information from the ESHA Research Database is also available.

This work presents *RecipeKG*, a publicly available knowledge graph comprising all recipes from the English Allrecipes.com[2] website, their ingredients, and corresponding nutritional information, social interactions metadata, and healthiness information calculated based on two international nutritional standards. Having this data in graph–based format has several advantages as follows. First, by linking concepts (e.g., ingredients) to external resources (e.g., [14]), the dataset can be enriched with knowledge (e.g., nutritional information of individual ingredients) that was not initially available. Such additional knowledge can facilitate research that is currently difficult (e.g., estimating calories from a recipe's ingredients), or outright impossible to conduct (e.g., learning to calculate the nutritional

[2]Allrecipes.com, a food–focused social network and recipe website, claims a traffic of 3 billion pages annually across 19 sites in 24 countries with recipes available in 13 languages [13].

Table 1

Comparison between *RecipeKG* and existing knowledge graphs.

	RecipeKG	FoodKG [14]	FoodWiki [17]	Open Food Facts [19]
Models recipes	**x**	x	x	
Models ingredients	**x**	x	x	x
Models nutritional data	**x**	x	x	x
Models intake guidelines	**x**			
Can be extended with additional guidelines	x			
Models (multiple) health scores	**x**			x
Can differentiate health scores by age and calorific intake	**x**			
Can be extended to include additional health scores	**x**			
Infers health scores	**x**			x
Taxonomy of categories	**x**	x		x
Includes social interaction metadata	**x**			

value of recipes). Second, edges in the knowledge graph explicitly represent relations between concepts (e.g., recipe ingredients), enabling analyses such as the one discussed in Section 4, as well as novel research directions (e.g., joint representation learning of recipes, ingredients, and nutrition facts). Third, by providing a consistent representation of nutritional data alongside diverse dietary guidelines, recipe healthiness tailored to ones age, gender, and daily calorific intake can be inferred. Forth, the data–driven evaluation of the healthiness of Internet recipes across categories, and their corresponding hierarchical structure, using widely accepted nutritional standards different dietary guidelines depending on user profiles in accordance to their age group and dietary intakes becomes possible.

The remainder of this paper is structured as follows: Section 2 reviews related work. Section 3 outlines the data collection and knowledge graph construction process. Section 4 presents a discussion pertaining to the healthiness of Allrecipes.com recipes. Finally, the potential impact of *RecipeKG* and potential future research directions are discussed in Sections 5 and 6, accordingly.

2. Related Work

Recipe–focused Knowledge Graphs. Despite the wealth of information on food recipes on the Web, only one recipe–focused knowledge graph, namely FoodKG [14], exists to date. FoodKG is complimentary to the knowledge graph presented here, since its recipes are sourced from the Recipe1M dataset [15]. FoodKG uses the Foodon Ontology [16] to categorize food, nutritional data about food, and chemical components of food. Similar to FoodKG, *RecipeKG* reuses some of the classes of Foodon Ontology [16]. However, unlike *RecipeKG*, FoodKG is oriented towards consumers who wish to determine what recipe to make based on ingredients at hand while accounting for constraints, such as allergies. It does not include nutritional information about food or recipes, nor does it facilitate the data–driven calculation of easily understandable health scores. Applications such as [17, 18] are narrow in scope (e.g., health, disease, or diet), and have been developed for a specific target audience. For instance, FoodWiki [17] is a mobile system that provides semantic–based suggestions about packaged food in the Turkish market using reasoning about side effects of and allergens in food based on properties of products captured using an Ontology and health condition of a given user. Similarly, the food and nutrition ontology described in [18] is used to help individuals identify recipes and ingredients, as well as monitor possible risk factors, aligned with hypertension, a chronic illness. Finally, Open Food Facts [19] is an open platform that allows individuals to add and edit information about packaged foods from around the world. Open Food Facts uses a scoring system that is similar to the FSA score used in this study. However, the focus is on packaged foods only. [20] provides a comprehensive review on food knowledge graphs. Table 1 summarizes the main differences between the proposed resource, i.e., *RecipeKG*, and most relevant, existing knowledge graphs.

Internet recipes' healthiness studies. We demonstrate the utility of *RecipeKG* by conducting a large–scale evaluation of Allrecipe.com recipes' healthiness. We therefore summarize works that have studied online recipes healthiness before [21–23]. Specifically, [21] conducted a small scale study (100 recipes for each category obtained from 3 supermarkets and 5 recipe books) to compare the nutritional content of main dishes from television chefs and ready meals sold in supermarkets. The authors of that seminal study are all public health experts that posed similar questions to the analysis in Section 4. Similar to this work, [21] used dietary guidelines from the World Health Organization [8], and the United Kingdom Food Standards Agency [24]. Follow up work [22, 23] included recipes from Allrecipes.com to the comparison, and found them to be unhealthy using FSA and WHO health criteria. Unlike prior research, this work offers insights into the healthiness of recipes on the Allrecipes.com website by emphasizing semantics in each step during the data collection and analysis process, and can therefore be used to provide explanations for each step involved in producing recipe recommendations.

3. Dataset Collection and Knowledge Graph Construction

The URLs of a total of $77,835$ recipes that were published on the main site of Allrecipes.com between the years 1997 and 2021 were crawled on December 14th, 2021 using its XML sitemap, and a standard crawler implemented in Python with Beautiful Soup. For every recipe (i) its title, (ii) average rating and number of reviews it has received, (iii) its classification, and (iv) the year of its publication, (v) the recommended number of servings, (vi) preparation and cooking times, and (vii) the complete list of ingredients and their corresponding quantities, as well as (viii) all available nutritional information about each recipe was extracted from linked data encoded within each recipe webpage as JSON–LD. The recommended number of servings (i.e., yield) provides some context for the nutritional information. Having the right recipe yield ensures that the correct % daily value is computed for each nutrient. The nutritional content for each recipe is reported by Allrecipes.com by matching the contained ingredients with those in the ESHA research database [25].

3.1. Data Preprocessing

The first step in processing the recipes is extracting relevant metadata encoded in JSON–LD in the corresponding recipe webpages. Although well structured, and adhering to the Schema.org schema, this metadata needed to be preprocessed before being ingested into the knowledge graph. We particularly faced two challenges with parsing ingredients and categories a recipe may belong to, as described below.

First, ingredients appear on the recipe metadata as strings (e.g., "1 cup brown sugar"). We processed each string using a conditional random field (CRF) model [26] in order to identify the ingredient itself (e.g., "brown sugar"), the quantity of the ingredient used in the recipe (e.g., "1"), and the unit of measurement (e.g., "cup"). The CRF model has been trained with $\sim 180,000$ human–labeled data from the New York Times cooking database, and is publicly available on GitHub[3]. We additionally applied a lemmatizer to promote homogeneity (i.e., prevent having multiple copies of the same unit appearing in different versions due for example to singular or plural form being used across recipes). We evaluated the accuracy of the CRF model by manually inspecting a total of 800 ingredients (i.e., 100 ingredients randomly sampled from each of the 8 sitemaps in our dataset, so as to increase coverage). Specifically, the output of the CRF model was manually labeled as correct (black), partially correct (amber), or erroneous (red) as shown in Table 2. Overall, the CRF model achieved 91.75% accuracy, with only 57 erroneous and 9 partially correct outputs out of the 800 samples.

Once ingredients were identified using the CRF model, an attempt was made to match each of them with ingredients in FoodOn [16], so that the corresponding, existing IRI could be used. A new IRI was constructed for each ingredient not matching an existing ingredient in the knowledge graph. Specifically, the `SequenceMatcher` function from the *difflib* library [27] was used to calculate the similarity between words appearing in the ingredient name (returned by the CRF model), and each ingredient in the knowledge base. All matches above a threshold of 70% were considered[4], and the IRI of the matching ingredient with the highest similarity score was used instead of the ingredient name string in the knowledge graph. In the event no match could be found, a new URI was created for the ingredient. To evaluate the accuracy of the matching process, we manually inspected a total of 729 ingredients from 80 recipes randomly sampled from each of the 8 sitemaps in our dataset. In all, we found this matching step to

[3] https://github.com/nytimes/ingredient-phrase-tagger

[4] This threshold was empirically found to produce the best results.

Table 2
Ingredient identification examples using CRF. Sample inputs to the CRF model are color coded to denote correctly identified ingredients (black), erroneous or misdetected ingredients (red), and borderline detections (amber).

Input string	Quantity	Unit	Name
2 tablespoon olive oil	2	tablespoon	olive oil
3/4 teaspoon curry powder	3/4	teaspoon	curry powder
1 large orange bell pepper, coarsely chopped	1	-	orange bell pepper
3 tablespoon crunchy peanut butter	3	tablespoon	butter
1/4 teaspoon red pepper flakes	1/4	teaspoon	-
3 egg yolks	3	-	egg
1 fluid ounce pineapple juice	1	ounce	fluid,pineapple juice
4 Granny Smith apples - peeled, cored and finely diced	4	-	apple
1 1/2 teaspoon minced lemon zest	1 1/2	teaspoon	lemon

be 100% accurate, as long as the CRF output was correct. In cases where the CRF model produced a partially correct or erroneous an ingredient name (e.g., "butter" as opposed to "peanut butter"), an erroneous IRI was assigned to that ingredient (e.g., the butter IRI would be assigned as opposed to the IRI referring to peanut butter). In the end, a total of 89 ingredients from FoodOn Ontology, and $6,220$ new ingredient IRIs, that were not part of the FoodOn Ontology, were included in *RecipeKG*.

Similar processing was performed to the nutritional information provided by Allrecipes.com for each recipe (see the overlayed window to the right in Fig. 2). Specifically, information about each nutrient appears in the order of quantity and unit. Once separated, these were linked to the corresponding recipe and stored in the knowledge graph.

Second, Allrecipes.com classifies recipes in numerous ways, including based on *ingredients* used (e.g., fruits and vegetables), a *region* (e.g., the midwest region of US), the *course* the dish is proposed for (e.g. appetizer or lunch), and any *holidays* the dish may be associated with (e.g, Thanksgiving). At the time of our crawl, there were 24 main categories on the Allrecipes.com website, and each recipe could be associated with one or more such categories. Categories are intuitively organized in a taxonomy (e.g., "Burger Recipes" is included under "Main Dishes", which we wished to reconstruct in order to be able to query *RecipeKG* for recipes in any given category. Thankfully, the categories a recipe belongs to are provided in a top down manner in its metadata. Therefore, we associated a recipe with the IRI of the lowest level category that appears in the recipe metadata, and created a subclass relationship between each subsequent pair of upper–level categories. For instance, the sample recipe in Fig. 2 would be associated with category "Date", which would then be identified as subcategory of "Fruit Cookie Recipes" that would in turn be stored as subcategory of "Cookies" and so on all the way to "Recipe Category" as shown in Fig. 3a. This way each recipe may be associated with multiple categories (or categorizations) that may themselves be a subcategory, establishing a path to the root (i.e., "Recipe category") of the taxonomy shown in Fig. 3b. Obviously, there could be issues, such as misspellings, inauthentic or misplaced recipes to a category. However, Allrecipes performs a moderation check for validity before a recipe is published.

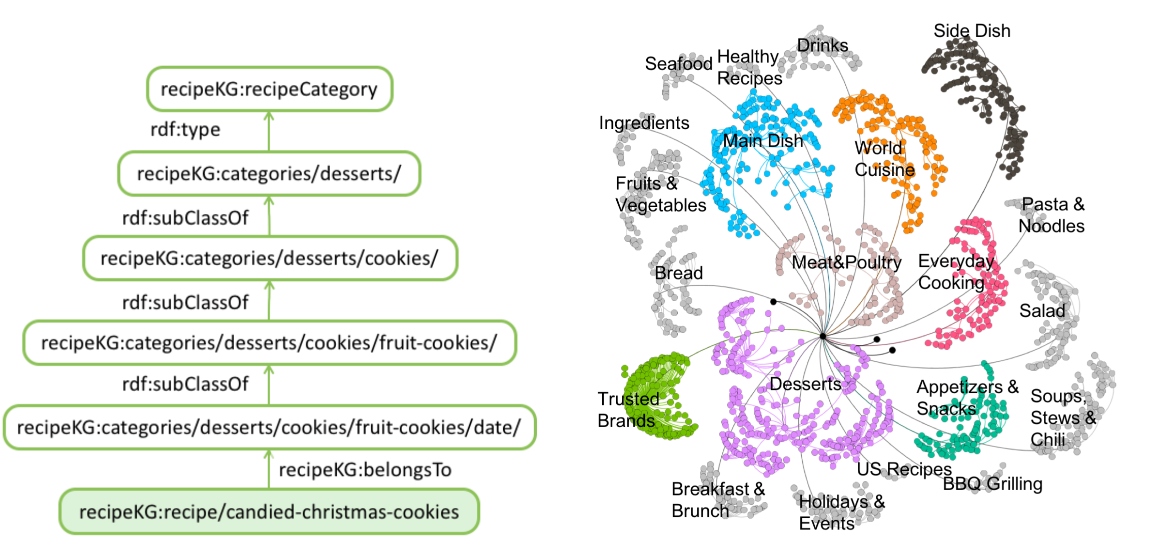
Finally, each recipe is assigned a Uniform Resource Identifier (URI) so that different recipes for the same thing (e.g., two different Christmas cake recipes by two individuals) can be distinguished, and their corresponding nutritional information can be retrieved, so that their healthiness can be evaluated.

3.2. Health Score Calculation

To estimate the healthiness of recipes in a data–driven manner, we considered two internationally recognized standards for measuring the healthiness of meals, namely the "Dietary Guidelines for Americans" by the United States Department of Agriculture (USDA) [28], and the "Guide to creating a front of pack (FoP) nutrition label" by the United Kingdom Food Standards Agency (FSA) [24]. Nevertheless, our ontology is extendable to other standards.

3.2.1. USDA Score.

The USDA defines daily ranges for macronutrients, minerals and vitamins. Following the approach of [22], which was conducted by public health experts, we focused on the 7 most important: (i) carbohydrates, (ii) protein, (iii) fat, (iv) saturated fat, (v) sugar, (vi) sodium, and (vii) fiber. For illustration purposes we use the corresponding ranges for



(a) (b)

Fig. 3. (a) Each recipe (an instance of schema:Recipe) in *RecipeKG* is associated with one or more categories, which in turn form a rich taxonomy, reconstructed from Allrecipe.com metadata. (b) Visualization of the reconstructed taxonomy. The black node in the middle corresponds to the root (i.e., "Recipe category").

females $19-30$ years of age to determine a USDA score. Nevertheless, our ontology allows ranges to be modeled for different age and daily intake scenarios. For each nutrient (e.g., calories), we derived the limit of content in grams by the given percentage of energy for that macronutrient. As an example for daily calorific intake of 2000, fat content must be between $\%20-\%35$. The minimum level of 20% corresponds to 400 calories, resulting in 44.4 grams of daily fat (1 gram fat $=9$ calories[5]). For sodium and fiber, density was calculated by dividing the calorie of recipe to fiber/sodium content and compared with the proposed density. The $\%$ daily value (i.e., how much a nutrient in a single serving of a recipe contributes to the daily recommended amount of a nutrient) was computed next. If a nutrient did not exceeded 20% of its daily value per serving[6], a recipe was awarded a point, for a total score between 0 and 7, with 0 meaning none of the ranges are fulfilled (totally unhealthy), and 7 meaning all ranges are met (very healthy). For protein and fiber, a recipe was awarded a point respectively, if the corresponding nutrient exceeded 20% of its daily value per serving, as high fiber and protein diets are considered to be beneficial.

3.2.2. Front of Pack Nutrition Label.

To help consumers make informed, and potentially healthier, choices when buying food, as well as to encourage food and drink companies to improve the nutritional quality of their products, many countries around the world have introduced front of pack nutrition labels [29]. Such labels strive to make nutrition information noticeable and easily understood. For example, the British Food Standards Agency uses a three colour–coded system (also known as the traffic–light labelling system) to indicate whether a food has high, medium or low amounts of fat, saturated fat, sugars and salt, as shown in Fig. 1b. Other notable systems include the French Nutri–Score, and the Australian Health Star Rating system [29]. *RecipeKG* can model any such score. For illustration purposes, we use FSA's "Traffic Light Labeling", because it is easier for people to understand, and it includes detailed information about nutrition and recipes, such as daily intakes, energy, and portion sizes. In order to derive a traffic light labeling score, we considered the same 4 macronutrients (i.e., sugar, sodium, fat and saturated fat) used in [22]. Different to that work, we assigned an integer value between 0 to 2 for each nutrient according to its color. Specifically, we

[5] https://www.nal.usda.gov/legacy/fnic/how-many-calories-are-one-gram-fat-carbohydrate-or-protein

[6] The U.S. Food and Drug Administration guidelines available at https://www.fda.gov/food/new-nutrition-facts-label/how-understand-and-use-nutrition-facts-label were used for this assessment.

assigned each nutrient 2 for green (low), 1 for amber (medium), 0 for red (high). By summing the score of individual nutrients, a score between 0 (very unhealthy) and 8 (very healthy) was derived for each recipe using a SWRL rule.

3.3. Representing Recipes in RDF

RecipeKG comprises a total of 1,641 classes and 49 properties to describe recipes, categories they belong to, ingredients, nutritional information, and health scores. To do so, *RecipeKG* primarily relies on Schema.org, that defines, maintains, and promotes structured data on the Internet, and provides a vocabulary for encoding recipe information in web pages [30]. For instance, social interactions metadata are represented using Schema's `bestRating`, `worstRating`, `reviewCount`, `ratingCount`, and `ratingValue` properties. Beyond Schema.org, *RecipeKG* reuses the `hasIngredient` relationship from the WhatToMake ontology [31] to link ingredients to each recipe. It also defines its own classes and properties (as shown in Fig. 4) to associate health scores to recipes.

An illustrative example is shown in Fig. 5a. In the example, the specific "Candied Christmas Cookies" recipe is linked to an ingredient, identified as an instance of a named class (i.e., `heals:BakingSoda`) in the WhatToMake ontology through the `wtm:hasIngredient` property. The quantity and unit of measurement used are also associated to the ingredient instance through the `recipeKG:quantity` and `recipeKG:unit` properties, respectively. Note that although normalizing the measurement units would be desirable, normalization is a challenging research problem in itself, which we plan to address in future work. Nevertheless, potential discrepancies between measurement units (e.g., cups versus grams) does not affect our study, since nutritional information for each recipe is recorded directly as calculated by Allrecipes.com.

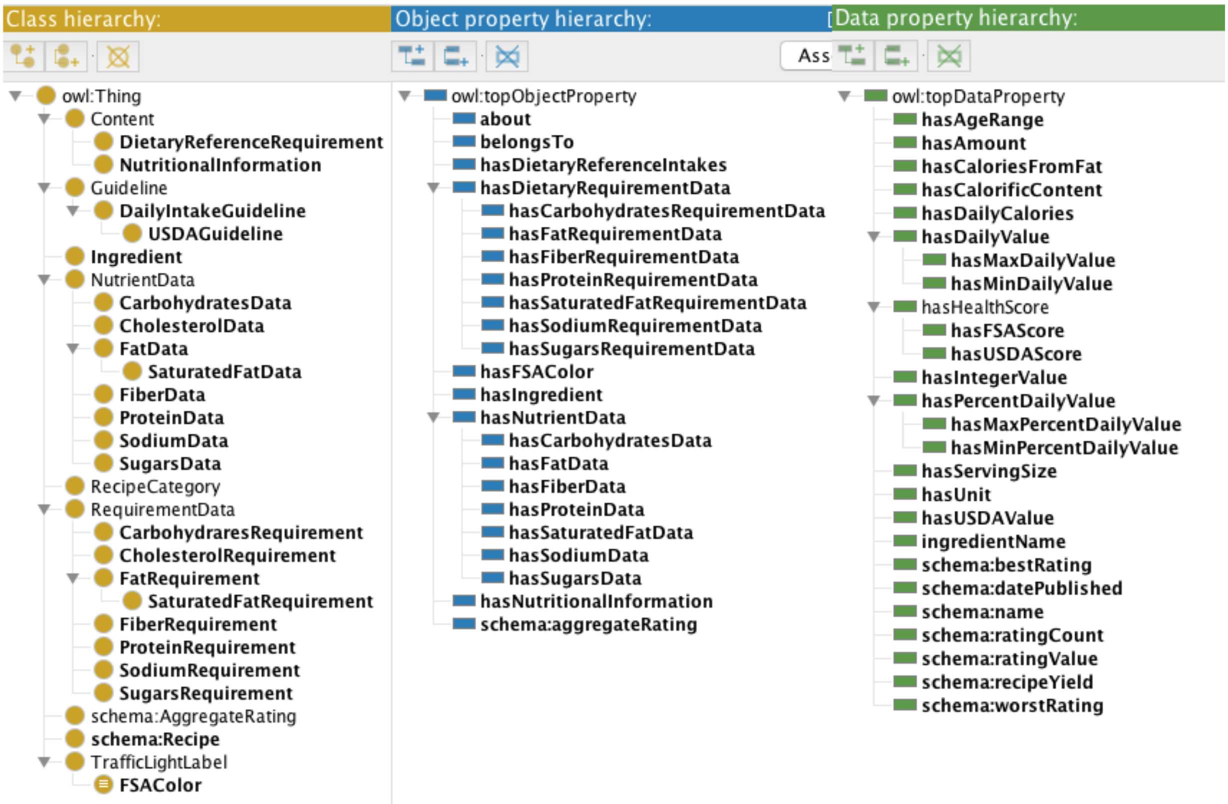
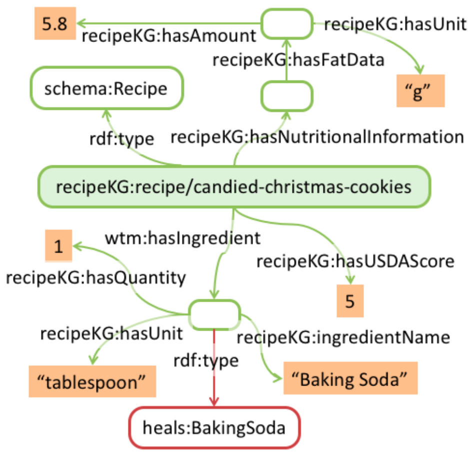


Fig. 4. List of classes (left), and object (center) and data properties (right) defined in the *RecipeKG* ontology. Note that categories, such as those shown in Fig. 3a, are subclasses of the "RecipeCategory" class, and are thus not listed here.



(a)

```
@prefix : <https//.../idiaslab/ontologies/recipeKG.owl#> .
@prefix wtm: <http://purl.org/heals/food/> .
@prefix schema: <http://schema.org/> .
...
:candiedchristmascookies rdf:type schema:Recipe ;
    schema:name "Candied Christmas Cookies" ;
    :hasServingSize 72 ;
    wtm:hasIngredient [rdf:type :egg; :hasQuantity "2"], ... ,
        [rdf:type :butter; :hasQuantity "1/2" ; :hasUnit "cup"],
        [rdf:type :milk; :hasQuantity "1" ; :hasUnit "tablespoon"],
        [rdf:type :candiedcherries; :hasQuantity "3" ; :hasUnit "cup"] ;
    :hasNutritionalInformation [rdf:type :NutritionalInformation ;
        :hasFatData [ rdf:type :FatData ;
            :hasAmount "5.8"^^xsd:float ; :hasUnit "g"^^xsd:string ;
            :hasFSAColor :FSAAmber ; :hasUSDAValue 1] ;
        :hasFiberData [ rdf:type :FiberData;
            :hasAmount "1.2"^^xsd:float ; :hasUnit "g"^^xsd:string ;
            :hasUSDAValue 0] ;
        :hasSodiumData [ rdf:type :SodiumData ;
            :hasAmount "33.6"^^xsd:float ; :hasUnit "mg"^^xsd:string ;
            :hasFSAColor :FSARed ; :hasUSDAValue 0] ;
        ...
        :hasCalorificContent "110.9"^^xsd:float ] ;
    :hasFSAScore 4 ; :hasUSDAScore 5 .
```

(b)

Fig. 5. (a) RDF representation of the candied-christmas-cookies recipe, its USDA healthscore, fat data, and one ingredient. Note that only a small subset of the recipe's metadata is depicted here. Note that datatypes, such as `recipeKG:hasQuantity` and `recipeKG:hasUSDAscore`, are modeled in the ontology. Both `:hasQuantity` and `:hasUSDAscore` are modeled as string literals, since the former can take many forms (e.g., fraction, integer, float) and the latter describes measurement units. At the same time, `reciprKG:hasUSDAscore` is modeled as `xsd:integer` $\in [0, 7]$. (b) Corresponding turtle representation.

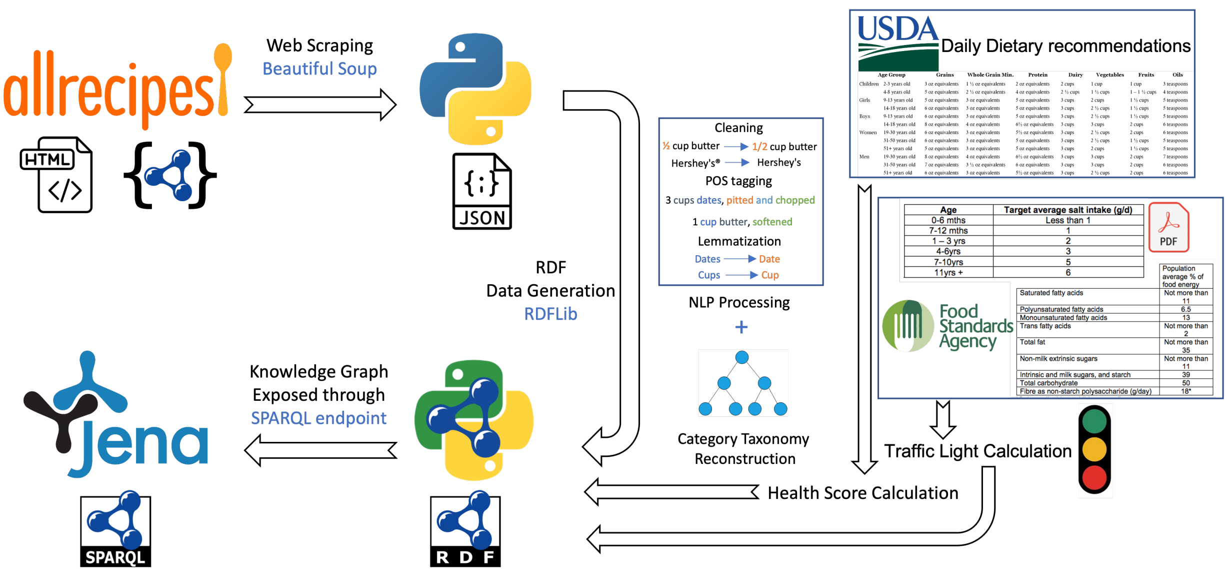


Fig. 6. Overview of the *RecipeKG* pipeline from data collection to data publication.

Fig. 6 shows the complete pipeline, that begins with the data assimilation from Allrecipes.com. After preprocessing (see Section 3.1), recipe metadata (in JSON-LD format) are translated to their corresponding RDF representation using the RDFLib [32] library in Python. Once in RDF form, the data is stored in an Apache Jena Fuseki SPARQL server and exposed through a publicly accessible SPARQL endpoint[7]. Two health scores are also com-

[7]https://recipekg.arcc.albany.edu

Table 3
Basic statistics of the RecipeKG knowledge graph.

Number of RDF triples	12, 944, 500
Number of recipes	77, 835
Number of ingredients	6309
Number of categories	1, 611
Number of SWRL rules	29
Number of nutrients	11

Table 4
Namespaces corresponding to prefixes

recipeKG	<http://purl.org/recipekg/>
wtm	<http://purl.org/heals/food/>
heals	<http://purl.org/heals/ingredient/>
rdfs	<http://www.w3.org/2000/01/rdf-schema#>
schema	<https://schema.org/>
xsd	<http://www.w3.org/2001/XMLSchema#>

puted and stored in the repository for each recipe as described in Section 3.2. Table 3 provides an overview of the basic statistics of *RecipeKG*, whereas Table 4 lists the namespaces and corresponding prefixes.

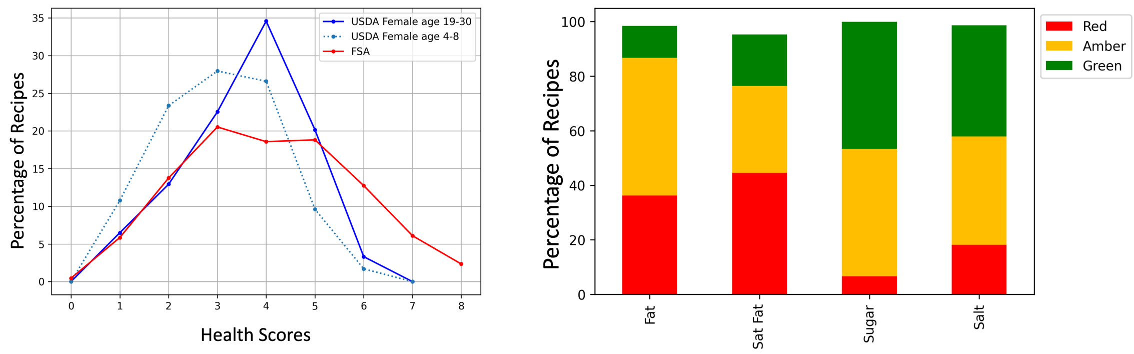
3.4. Accessibility and Sustainability

We make *RecipeKG* data findable, accessible, interoperable, and reusable[8] by (i) making the RDF data retrievable through an open and free, globally unique and eternally persistent identifier at https://doi.org/10.7910/DVN/99PNJ5 provided by Dataverse[9], (ii) including rich citation terms[10], compliant with Dublin Core[11], and (iii) opening it to the public domain using a CC0 – "Public Domain Dedication" license. The RecipeKG ontology is available at http://purl.org/recipekg/. We additionally make our source code, along with documentation, and dataset (both in JSON-LD and RDF format) publicly available at https://github.com/IDIASLab/RecipeKG. To encourage the collaborative and sustainable extension of RecipeKG Ontology, we monitor the repository's issue tracker for requests of new terms/classes, as well as for reported errors or specific concerns related to the ontology.

4. Analyzing the Healthiness of Allrecipes Recipes

Inspired by the seminal study of recipes' healthiness, [22], which was conducted by public health experts, we illustrate the value of *RecipeKG*, by showcasing its ability to augment the workflow of researchers studying the healthiness of online recipes. We begin by plotting the distribution of recipes by number of fulfilled criteria according to the two health scores described in Section 3.2. The result, shown in Fig. 7a, indicates that only 10 (%0.012) recipes fail all USDA guidelines (very unhealthy), whereas no recipe meets all USDA criteria (very healthy). The rest of the recipes meet $1-6$ criteria, with the majority (%77.25) meeting $3-5$ guidelines. Note that this analysis is for females between $19-30$ years of age on a 2,000 calories per day diet. When adjusted for a different demographic (e.g., females between $4-8$ years old on a 1,200 calorie per day diet), the distribution shifts to the left (i.e, less criteria are met, for instance, due to reduced level of calorific intake per day), indicating less healthy recipes for this age group. The comparison of recipes by their healthiness according to different dietary guidelines and by user profile is only possible because of the interlinking of concepts in *RecipeKG*.

[8] https://force11.org/info/the-fair-data-principles/
[9] https://dataverse.org/
[10] http://dublincore.org/documents/dcmi-terms/
[11] https://www.dublincore.org/



(a) (b)

Fig. 7. (a) Distribution of recipes by number of fulfilled criteria according to USDA health score. (b) FSA traffic light assessment of the same recipes.

With respect to FSA score, 360 (%0.46) recipes are labeled red (i.e, very unhealthy), as compared to $1,823$ (%2.34) being labeled green (i.e, very healthy). Fig. 7b further indicates that only a small percentage of recipes are labeled green with respect to fat and saturated fat. Specifically for fat, more than half the recipes are amber, and for saturated fat, almost half of the recipes are red, indicating that many, if not most, recipes constitute unhealthy choices with respect to these nutrients.

To better understand the reasons behind this finding, we computed the average nutritional properties of recipes across 16 main categories from Allrecipes.com, each associated with more than 100 recipes, as shown in Table 5. The nutritional properties of recipes in categories with less than 100 recipes each, are summarized as "Other" in this table. The last row of Table 5 indicates that overall, recipes are not as healthy as one would expect, with an average USDA and FSA score of 3.79 and 4.27, accordingly. Particularly worrisome are the average values of fat and saturated fat (both red according to FSA traffic lighting system), and sodium content (amber). A similarity can be observed between the average USDA and FSA scores across categories, with the average healthiness of recipes in different categories fluctuating only slightly.

"Seafood" and "Bread" recipes are the healthiest, while "Desserts" and "Main Dish" recipes or recipes involving pasta and meat are the least healthy. Recipes in these categories exhibit, on average, high calorific content, fat, and saturated fat, and cholesterol, and protein. "Drink" recipes seem to fair comparatively well, although they exhibit a high sugar average (second only to recipes in the "Desserts" category). Less intuitive results include that recipes in the "Fruit and Vegetables" category are low on fiber content and high on fat. This finding contradicts the study in [23]. However, this discrepancy is in fact quite easy to explain, as it directly relates to the way recipes are organized into categories by Allrecipes.com, and consequently in *RecipeKG*. In reconstructing the category taxonomy (c.f. Section 3.1), we have unambiguously linked recipes to categories. This results in a more accurate representation of the corpus, even if it means including in the "Fruit and Vegetables" category recipes that do not only contain fruits and vegetables as ingredients, but additionally feature other, often unhealthy, ingredients (e.g., sugar and butter in high calorie desserts, and meat in stuffed vegetable recipes).

By additionally considering social interactions metadata (i.e., number of ratings and average rating score), we observe that popular categories exhibit a high number of ratings per recipe, but also that categories with high average health score tend to have higher average rating than categories with less healthy recipes. For instance, both "Pasta & Noodles" (one of the least healthy categories) and "Bread" recipes (the category with the highest average health score) have a high number of ratings per recipe, however, "Bread" recipes have a higher average rating than the recipes in the less healthy "Pasta & Noodles" category.

Table 5

Average nutritional content and health score per recipe by category. The most healthy (green) and least healthy (red) category is identified for each nutrient. “Drinks” are excluded to avoid biasing the analysis, whereas “Sodium” and “Cholesterol” are omitted for the Table to fit. The number of recipes in each category, and USDA and FSA scores are denoted by n, U, and F accordingly.

Category	n	U	F	**Cal. (kCal)**	**Fat (g)**	**Sat. Fat (g)**	**Carb. (g)**	**Prot. (g)**	**Fib. (mg)**	**Sug. (g)**	**Rating Count**	**Avg. Rating**
Appetizers&snacks	5930	4.06	4.81	**217.28**	13.82	5.04	**15.69**	8.69	1.87	**4.71**	66.40	4.34
Bread	3533	3.64	**4.99**	237.56	**9.47**	**3.52**	33.98	5.19	1.91	12.15	137.82	4.30
Breakfast&brunch	3759	3.62	4.00	331.91	17.01	6.88	33.23	13.09	3.12	12.68	100.86	4.32
Desserts	14163	**2.62**	3.67	302.30	15.11	7.08	39.87	**4.04**	**1.64**	**26.06**	94.03	4.24
Drinks	2837	3.26	4.90	237.03	3.93	3.09	32.61	2.46	1.43	25.40	26.26	4.37
Everyday cooking	1572	3.88	4.23	339.92	17.17	6.41	33.00	15.90	4.07	8.83	65.74	4.27
Fruits&vegetables	1155	3.31	3.98	303.52	15.49	5.93	34.64	8.71	3.81	15.83	82.83	4.31
Main dish	6475	3.76	**3.26**	463.17	23.73	9.12	37.25	25.37	3.41	7.61	130.58	4.31
Meat&poultry	5412	3.99	3.40	460.03	**24.53**	8.64	27.28	**31.61**	2.46	8.64	120.78	4.30
Pasta&noodles	388	3.53	3.39	**469.37**	21.89	**9.24**	**47.62**	21.31	4.10	7.46	143.87	4.18
Salad	4101	3.86	4.36	303.11	18.09	4.05	26.64	10.63	4.06	9.23	61.92	4.39
Seafood	1363	**4.31**	3.97	411.61	20.68	7.47	26.27	29.48	2.51	4.83	95.10	4.32
Side dish	9295	3.78	4.71	221.08	12.24	5.10	24.08	6.17	3.04	8.32	77.98	4.35
Soups stews&Chili	5495	4.17	3.96	338.00	15.55	6.26	31.99	18.44	**5.77**	6.70	101.91	4.35
Trusted brand recipes	3667	3.66	3.91	371.78	18.50	7.26	35.49	16.90	3.19	10.56	28.25	4.41
World cuisine	7558	3.70	3.67	407.14	20.33	7.85	35.44	20.71	3.38	11.03	100.09	4.32
Other	$\leqslant 100$	4.03	4.79	204.08	11.04	4.35	18.54	9.06	1.88	8.07	90.51	4.56
All	76,906	3.79	4.27	301.79	15.17	5.87	28.47	13.32	2.78	10.38	85.77	4.18

5. Potential Impact and Reusability

To the best of our knowledge, *RecipeKG* is the first recipe–focused knowledge graph that integrates recipe and their ingredients with social interactions metadata, all while enabling inference of recipe healthiness using multiple nutritional standards. Beyond the healthiness evaluation of online recipes presented in Section 4, *RecipeKG* has the potential to catalyze new research. For instace, *RecipeKG* makes it possible to assess the accuracy of nutritional information in online recipes by linking ingredients to corresponding entities in FoodKG [14], which in turn links ingredients to entities in the FoodOn ontology [16] and the USDA's nutrition data[12]. Such linking will enable a direct calculation of a recipe's nutritional content (i.e., calories and % daily value of nutrients) and its corresponding health score, and will allow a comparison between the semantically computed nutritional content and the “ground-truth” available on websites such as Allrecipes.com based on the ESHA research database [25].

Another potential application of *RecipeKG* is improving the functionality of recipe sharing websites such as Allrecipes.com, where users can search for recipes containing certain types of ingredients. Currently, recipe sharing websites such as Allrecipes.com, cannot identify recipes appropriate for certain health conditions (e.g., high blood pressure) that place restrictions on nutritional intake. Since the health score of recipes and information about ingredients is encoded directly within the knowledge graph, *RecipeKG* can be used to assist with such informational tasks as demonstrated by the SPARQL query listed in the Fig. 8a. Similarly, the SPARQL query listed in the Fig. 8b can be used to identify recipes belonging to a specific category and involving an ingredient of choice, while at the same ensuring their sugar content is low. Such queries can be particularly useful to users with certain health conditions (e.g., diabetes) that place restrictions on nutritional intake.

Each of the queries described above can be answered by querying *RecipeKG* using SPARQL, since information like the relationships between recipes and categories is encoded directly within the knowledge graph. Specifically, since *RecipeKG* structures its knowledge of recipe categories in a hierarchical way, it can determine whether recipes belong to a certain category even if they are not directly linked to such category. For instance, recipes listed under *recipeKG:cate gories/main-dish/casseroles/*, *recipeKG:categories/main-dish/pork/pork-chops/ boneless/*, and *recipeKG:categories/main-dish/pork/pork-chops/boneless/* would all be retrieved using the query listed in Fig. 8b, even if they are all associated with subcategories of the “Main Dish” category. Finally, the third SPARQL query

[12] https://catalog.data.gov/dataset/food-and-nutrient-database-for-dietary-studies-fndds

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX schema: <https://schema.org/>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX recipeKG:<http://purl.org/recipekg/>
PREFIX categories:<http://purl.org/recipekg/categories/>
PREFIX ingredient: <http://purl.org/heals/ingredient/>
PREFIX food: <http://purl.org/heals/food/>

SELECT ?recipe ?fsa ?w
WHERE {
  ?recipe recipeKG:hasFSAScore ?fsa.
  ?recipe recipeKG:hasNutritionalInformation ?ni.
  ?ni recipeKG:hasProteincontent ?p.
  ?p recipeKG:hasAmount ?a.
  FILTER (?fsa >=6 && ?a >= 25)
}
```

(a)

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX schema: <https://schema.org/>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX recipeKG:<http://purl.org/recipekg/>
PREFIX categories:<http://purl.org/recipekg/categories/>
PREFIX ingredient: <http://purl.org/heals/ingredient/>
PREFIX food: <http://purl.org/heals/food/>

SELECT ?recipe
WHERE {
  ?recipe recipeKG:belongsTo ?category .
  ?category rdfs:subClassOf* <http://purl.org/recipekg/categories/main-dish/>.
  ?recipe food:hasIngredient ?x.
  ?x a  ingredient:Onion.
  ?recipe recipeKG:hasNutritionalInformation ?y.
  ?y recipeKG:hasSugarcontent ?z.
  ?z recipeKG:hasFSAColor recipeKG:FSAGreen.
}
```

(b)

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX schema: <https://schema.org/>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX recipeKG:<http://purl.org/recipekg/>
PREFIX categories:<http://purl.org/recipekg/categories/>
PREFIX ingredient: <http://purl.org/heals/ingredient/>
PREFIX food: <http://purl.org/heals/food/>

SELECT ?recipe
WHERE {
  ?recipe recipeKG:hasNutritionalInformation ?y.
  ?y recipeKG:hasSodiumcontent ?z.
  ?z recipeKG:hasFSAColor recipeKG:FSAGreen.
  ?z recipeKG:hasUSDAValue 0.
}
```

(c)

```
schema:Recipe(?r) ^ :hasNutritionalInformation(?r,
?i) ^ :hasFatData(?i, ?f) ^ :hasAmount(?f, ?a) ^
:hasDietaryReferenceIntakes(:USDA_Females19-30,
:USDA_Females19-30_DietaryReferenceIntakes) ^
:hasFatRequirementData(:USDA_Females19-
30_DietaryReferenceIntakes, :USDA_Females19-
30_FatDRI) ^ :hasMinDailyValue(:USDA_Females19-
30_FatDRI, ?fdv) ^ swrlb:multiply(?fp, ?fdv, 0.2) ^
swrlb:greaterThan(?a, ?fp) -> :hasUSDAValue(?f, 0)
```

(d)

Fig. 8. SPARQL queries to retrieve recipes that (a) have both high FSA score and protein content, (b) belong to a certain category, contain a given ingredient, and meet a certain nutritional constraint, and (c) exhibit a discrepancy between standards for a given nutrient. (d) Smple SWRL rule, used to determine whether or not to award a recipe a point for its fat content, given a specific dietary reference intake guideline. Rules are used to derive the FSA and USDA scores for each recipe. In addition, rules for each <age, calorific intake> tuple (not shown here) are needed for personalized USDA score calculation.

listed in Fig. 8c would retrieve those recipes whose sodium content would be considered healthy according to FSA but not according to USDA recommendations. Such functionality may be particularly valuable to researchers studying the healthiness of online recipes according to different criteria and can be retrieved because of the inferencing capabilities of *RecipeKG*.

Finally, consider the task of determining how to improve the nutritional contents of a recipe by substituting ingredients (e.g., replacing high–carb ingredients with low–carb alternatives) while maintaining the essence of the original recipe. By being interoperable with FoodOn ontology [16] and the USDA's nutrition information, *RecipeKG* can help identify which "unhealthy" ingredients to remove from a recipe as well as which substitution options are "healthy", all while keeping the healthscore of a recipe up to date. In this scenario the quantity and unit information recorded for each ingredient in *RecipeKG* can be used to compute the nutritional content of each ingredient in a recipe using USDA's food data, while taking into consideration that USDA presents nutritional information (such as grams of carbohydrates or milligrams of sodium) per 100g of each ingredient.

Beyond the potential applications listed above, we expect the new resource described in this work to spur significant followup work, considering the recent flair of papers related to food [6, 33–35], recipes and ingredients [36–38], and health (e.g., [14, 39–43]).

6. Conclusion

The Web is abundant with recipes contributed in recipe sharing websites. By transforming recipe data spanning a time period of 25 years from Allrecipes.com and their corresponding metadata into a knowledge graph, and augmenting such data with nutritional standards, a wide range of useful applications are made possible. *RecipeKG* is interoperable with and can be further linked to food–related ontologies (e.g., FoodOn [16]) and knowledge graphs (e.g., FoodKG [14]). Even though the analysis presented in Section 4 is restricted to Allrecipes.com, extending *RecipeKG* with the ability to store metadata about recipes from other websites (e.g., epicurious.com or simplyrecipes.com) or datasets (e.g., the Recipe1M dataset [15]) is straightforward. The knowledge graph construction process itself, although tailored to Allrecipes.com, can be augmented to facilitate integration with data from other recipe sharing websites that publish structured metadata. Similarly, the choice of the two health scores used for illustration purposes was driven by precedence in the literature. However, *RecipeKG* can easily accommodate other international standards for recipe health assessment. Finally, our focus was on individual recipes, although a healthy diet is created by combining a variety of food types. Combining recipes to recommend a daily meal plan that meets guidelines from official health organizations may require integration with other Ontologies, such as HeLiS [40].

As a new resource, the potential impact of *RecipeKG* can only be speculated. However, given the increasing research interest in healthy food recommendations, we anticipate this knowledge graph to spur new research in this domain. Future work will include a front–end to allow non–technical people interact with this resource, and learning embeddings of entities within the knowledge graph. We additionally plan to evolve and improve the RecipeKG ontology as our team identifies additional nutritional standards.

References

[1] P.K. Powell, J. Durham and S. Lawler, Food choices of young adults in the United States of America: a scoping review, *Advances in Nutrition* **10**(3) (2019), 479–488.

[2] A. Steptoe, T.M. Pollard and J. Wardle, Development of a measure of the motives underlying the selection of food: the food choice questionnaire, *Appetite* **25**(3) (1995), 267–284.

[3] J.C. Kozup, E.H. Creyer and S. Burton, Making healthful food choices: the influence of health claims and nutrition information on consumers' evaluations of packaged food products and restaurant menu items, *Journal of Marketing* **67**(2) (2003), 19–34.

[4] J. Barreiro-Hurlé, A. Gracia and T. De-Magistris, Does nutrition information on food products lead to healthier food choices?, *Food Policy* **35**(3) (2010), 221–229.

[5] E. Gorbonos, Y. Liu and C.T. Hoàng, NutRec: Nutrition oriented online recipe recommender, in: *2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*, IEEE, 2018, pp. 25–32.

[6] Y. Chen, A. Subburathinam, C.-H. Chen and M.J. Zaki, Personalized food recommendation as constrained question answering over a large-scale food knowledge graph, in: *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*, 2021, pp. 544–552.

[7] S. Shirai, O. Seneviratne, C. Chen, D. Gruen and D.L. McGuinness, Healthy Food Recommendation and Explanation Generation using a Semantically-Enabled Framework, in: *Proceedings of the ISWC 2021 Posters, Demos and Industry Tracks: From Novel Ideas to Industrial Practice co-located with 20th International Semantic Web Conference (ISWC 2021), Virtual Conference, October 24-28, 2021*, Vol. 2980, CEUR-WS.org, 2021.

[8] J.W.E. Consultation, Diet, nutrition and the prevention of chronic diseases, *World Health Organ Tech Rep Ser* **916**(i–viii) (2003), 1–149.

[9] A. Mente, L. de Koning, H.S. Shannon and S.S. Anand, A systematic review of the evidence supporting a causal link between dietary factors and coronary heart disease, *Archives of internal medicine* **169**(7) (2009), 659–669.

[10] G.J. Hankey, Nutrition and the risk of stroke, *The Lancet Neurology* **11**(1) (2012), 66–81.

[11] P. Greenwald, Role of dietary fat in the causation of breast cancer: point, *Cancer Epidemiology and Prevention Biomarkers* **8**(1) (1999), 3–7.

[12] J. Firth, J.E. Gangwisch, A. Borsini, R.E. Wootton and E.A. Mayer, Food and mood: how do diet and nutrition affect mental wellbeing?, *Bmj* **369** (2020).

[13] AELIEVE, Best Recipe Websites, 2021. https://aelieve.com/rankings/websites/category/food-beverage/best-recipe-websites/.

[14] S. Haussmann, O.W. Seneviratne, Y. Chen, Y. Ne'eman, J. Codella, C.-H. Chen, D.L. McGuinness and M.J. Zaki, FoodKG: A Semantics-Driven Knowledge Graph for Food Recommendation, in: *SEMWEB*, 2019.

[15] J. Marin, A. Biswas, F. Ofli, N. Hynes, A. Salvador, Y. Aytar, I. Weber and A. Torralba, Recipe1m+: A dataset for learning cross-modal embeddings for cooking recipes and food images, *IEEE transactions on pattern analysis and machine intelligence* **43**(1) (2019), 187–203.

[16] D.M. Dooley, E.J. Griffiths, G.P.S. Gosal, P.L. Buttigieg, R. Hoehndorf, M. Lange, L.M. Schriml, F.S.L. Brinkman and W.W.L. Hsiao, FoodOn: a harmonized food ontology to increase global food traceability, quality control and data integration, *NPJ Science of Food* **2** (2018).

[17] D. Çelik, FoodWiki: Ontology-Driven Mobile Safe Food Consumption System, *The Scientific World Journal* **2015** (2015).
[18] J. Clunis, Designing an ontology for managing the diets of hypertensive individuals, *International Journal on Digital Libraries* **20** (2018), 269–284.
[19] O.F. Facts, A collaborative, free and open database of food products from around the world, 2021. https://world.openfoodfacts.org.
[20] W. Min, C. Liu and S. Jiang, Towards Building a Food Knowledge Graph for Internet of Food, *arXiv e-prints* (2021), arXiv–2107.
[21] S.J. Howard, J. Adams and M. White, Nutritional content of supermarket ready meals and recipes by television chefs in the United Kingdom: cross sectional study, *The BMJ* **345** (2012).
[22] C. Trattner, D. Elsweiler and S.J. Howard, Estimating the Healthiness of Internet Recipes: A Cross-sectional Study, *Frontiers in Public Health* **5** (2017).
[23] C. Trattner and D. Elsweiler, Investigating the healthiness of internet-sourced recipes: implications for meal planning and recommender systems, in: *Proceedings of the 26th international conference on world wide web*, 2017, pp. 489–498.
[24] U. FSA, Guide to creating a front of pack (FoP) nutrition label for pre-packed products sold through retail outlets, *Food Standards Agency* (2014).
[25] ESHA, Nutrition Labeling Software, 2021. https://esha.com.
[26] C. Sutton, A. McCallum et al., An introduction to conditional random fields, *Foundations and Trends® in Machine Learning* **4**(4) (2012), 267–373.
[27] P.S. Foundation, difflib Library, 2022. https://docs.python.org/3/library/difflib.html.
[28] U.S.D.G.A. Committee, *Dietary Guidelines for Americans,2020-2025*, Vol. 9th Edition, U.S. Department of Agriculture and U.S. Department of Health and Human Services, 2020. https://dietaryguidelines.gov.
[29] A. Jones, B. Neal, B. Reeve, C.N. Mhurchu and A.M. Thow, Front-of-pack nutrition labelling to promote healthier diets: current practice and opportunities to strengthen regulation worldwide, *BMJ Global Health* **4**(6) (2019), e001882.
[30] Schema.org, , 2021. https://schema.org/.
[31] M. Qi, Y. Neeman, F. Eckhardt and K. Blissett, WhatToMake: A Semantic Web Application for Recipe Recommendation, *Rensselaer Polytechnic Institute* (2018).
[32] RDFlib, A volunteer-maintained collection of Python tools for working with RDF data, 2005. https://github.com/RDFLib/rdflib.
[33] M. Kolchin, A. Chistyakov, M. Lapaev and R. Khaydarova, FOODpedia: Russian food products as a linked data dataset, in: *European Semantic Web Conference*, Springer, 2015, pp. 87–90.
[34] J. Youn, T. Naravane and I. Tagkopoulos, Using word embeddings to learn a better food ontology, *Frontiers in artificial intelligence* (2020), 93.
[35] D. Park, K. Kim, S. Kim, M. Spranger and J. Kang, FlavorGraph: a large-scale food-chemical graph for generating food representations and recommending food pairings, *Scientific reports* **11**(1) (2021), 1–13.
[36] D. Li and M.J. Zaki, Reciptor: An effective pretrained model for recipe representation learning, in: *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2020, pp. 1719–1727.
[37] S.S. Shirai, O. Seneviratne, M.E. Gordon, C.-H. Chen and D.L. McGuinness, Identifying ingredient substitutions using a knowledge graph of food, *Frontiers in Artificial Intelligence* (2021), 111.
[38] Y. Tian, C. Zhang, R. Metoyer and N.V. Chawla, Recipe representation learning with networks, in: *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, 2021, pp. 1824–1833.
[39] M. Chokr and S. Elbassuoni, Calories prediction from food images, in: *Twenty-Ninth IAAI Conference*, 2017.
[40] M. Dragoni, T. Bailoni, R. Maimone and C. Eccher, Helis: An ontology for supporting healthy lifestyles, in: *International semantic web conference*, Springer, 2018, pp. 53–69.
[41] H. Jiang, W. Wang, M. Liu, L. Nie, L.-Y. Duan and C. Xu, Market2dish: A health-aware food recommendation system, in: *Proceedings of the 27th ACM International Conference on Multimedia*, 2019, pp. 2188–2190.
[42] J. Harashima, M. Hiramatsu and S. Sanjo, Calorie estimation in a real-world recipe service, in: *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34, 2020, pp. 13306–13313.
[43] R. Azzi, S. Despres and G. Diallo, NutriSem: A Semantics-Driven Approach to Calculating Nutritional Value of Recipes, in: *World Conference on Information Systems and Technologies*, Springer, 2020, pp. 191–201.
[44] B. Wansink, *Mindless eating: Why we eat more than we think*, Bantam, 2007.
[45] M. Kolchin and D. Zamula, Food Ontology: Ontology for Describing Food Products, 2013.
[46] M.N.K. Boulos, A. Yassine, S. Shirmohammadi, C.S. Namahoot and M. Brückner, Towards an "Internet of Food": Food Ontologies for the Internet of Things, *Future Internet* **7** (2015), 372–392.
[47] X. Chen, S. Jia and Y. Xiang, A review: Knowledge reasoning over knowledge graph, *Expert Systems with Applications* **141** (2020), 112948.
[48] M. Scrocca, M. Comerio, A. Carenini and I. Celino, Turning transport data to comply with EU standards while enabling a multimodal transport knowledge graph, in: *International Semantic Web Conference*, Springer, 2020, pp. 411–429.
[49] X. Wang, X. He, Y. Cao, M. Liu and T.-S. Chua, Kgat: Knowledge graph attention network for recommendation, in: *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, 2019, pp. 950–958.
[50] X. Zou, A survey on application of knowledge graph, in: *Journal of Physics: Conference Series*, Vol. 1487, IOP Publishing, 2020, p. 012016.
[51] A. Hogan, E. Blomqvist, M. Cochez, C. D'amato, G.D. Melo, C. Gutierrez, S. Kirrane, J.E.L. Gayo, R. Navigli, S. Neumaier, A.-C.N. Ngomo, A. Polleres, S.M. Rashid, A. Rula, L. Schmelzeisen, J. Sequeda, S. Staab and A. Zimmermann, Knowledge Graphs, *ACM Computing Surveys* **54**(4) (2022), 1–37–. doi:10.1145/3447772.

[52] J. Cardoso and A. Sheth, The Semantic Web and its applications, in: *Semantic Web Services, Processes and Applications*, Springer, 2006, pp. 3–33.