



**HAL**  
open science

# HUMMUS: A Linked, Healthiness-Aware, User-centered and Argument-Enabling Recipe Data Set for Recommendation

Felix Bölz, Diana Nurbakova, Sylvie Calabretto, Lionel Brunie, Armin Gerl,  
Harald Kosch

## ► To cite this version:

Felix Bölz, Diana Nurbakova, Sylvie Calabretto, Lionel Brunie, Armin Gerl, et al.. HUMMUS: A Linked, Healthiness-Aware, User-centered and Argument-Enabling Recipe Data Set for Recommendation. 17th ACM Conference on Recommender Systems (RECSYS23), Sep 2023, New York, NY, United States. 10.1145/3604915.3609491 . hal-04220182

**HAL Id: hal-04220182**

**<https://hal.science/hal-04220182>**

Submitted on 29 Sep 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



# HUMMUS: A Linked, Healthiness-Aware, User-centered and Argument-Enabling Recipe Data Set for Recommendation

Felix Bölz  
Felix.Boelz@uni-passau.de  
Univ Lyon, INSA Lyon, CNRS, UCBL,  
LIRIS, UMR5205  
Villeurbanne, France  
University of Passau  
Passau, Germany

Diana Nurbakova  
Sylvie Calabretto  
Lionel Brunie  
diana.nurbakova@insa-lyon.fr  
sylvie.calabretto@insa-lyon.fr  
lionel.brunie@insa-lyon.fr  
Univ Lyon, INSA Lyon, CNRS, UCBL,  
LIRIS, UMR5205  
Villeurbanne, France

Armin Gerl  
Harald Kosch  
Armin.Gerl@uni-passau.de  
Harald.Kosch@uni-passau.de  
University of Passau  
Passau, Germany

## ABSTRACT

The overweight and obesity rate is increasing for decades worldwide. Healthy nutrition is, besides education and physical activity, one of the various keys to tackle this issue. In an effort to increase the availability of digital, healthy recommendations, the scientific area of food recommendation extends its focus from the accuracy of the recommendations to beyond-accuracy goals like transparency and healthiness. To address this issue a data basis is required, which in the ideal case encompasses user-item interactions like ratings and reviews, food-related information such as recipe details, nutritional data, and in the best case additional data which describes the food items and their relations semantically. Though several recipe recommendation data sets exist, to the best of our knowledge, a holistic large-scale healthiness-aware and connected data sets have not been made available yet. The lack of such data could partially explain the poor popularity of the topic of healthy food recommendation when compared to the domain of movie recommendation. In this paper, we show that taking into account only user-item interactions is not sufficient for a recommendation. To close this gap, we propose a connected data set called *HUMMUS* (Health-aware User-centered recoMMendation and argUment-enabling data Set) collected from *Food.com* containing multiple features including rich nutrient information, text reviews, and ratings, enriched by the authors with extra features such as Nutri-scores and connections to semantic data like the *FoodKG* and the *FoodOn* ontology. We hope that these data will contribute to the healthy food recommendation domain.

## CCS CONCEPTS

• **Computing methodologies** → **Learning from implicit feedback**; • **Information systems** → **Data cleaning; Recommender systems**; • **Applied computing** → **Health informatics**.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

*RecSys '23, September 18–22, 2023, Singapore, Singapore*

© 2023 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 979-8-4007-0241-9/23/09...\$15.00

<https://doi.org/10.1145/3604915.3609491>

## KEYWORDS

Recipe data set, Knowledge graph, Nutrition scores, Healthiness-aware recommendation, Explainable recommendation

### ACM Reference Format:

Felix Bölz, Diana Nurbakova, Sylvie Calabretto, Lionel Brunie, Armin Gerl, and Harald Kosch. 2023. HUMMUS: A Linked, Healthiness-Aware, User-centered and Argument-Enabling Recipe Data Set for Recommendation. In *Seventeenth ACM Conference on Recommender Systems (RecSys '23), September 18–22, 2023, Singapore, Singapore*. ACM, New York, NY, USA, 11 pages. <https://doi.org/10.1145/3604915.3609491>

## 1 INTRODUCTION

In today's world, there are 1.9 billion overweight people and 650 million of those suffering from obesity. Unfortunately, the trend is upward, as global obesity rates nearly tripled between 1975 and 2016. Possible consequences could range from a moderately reduced health level up to avoidable premature deaths [60]. Those high numbers also have an impact on society (for example, health insurances), and thus the general public has to bear immense costs as a result [38]. Among other things, this influences the urgency of how the issue of nutrition is perceived in society, politics and science. The reason is that healthy nutrition plays, besides physical activity, a key role in the prevention of overweight and obesity.

We focus in our work on how interactive recommendation systems can contribute to counteracting the current trend and encourage individuals to adopt healthier eating habits. Unfortunately, it is notable that the field of healthy dietary recommendation systems does not receive nearly as much attention as other domains of recommendation systems such as those in the fields of movie and product recommendation [49].

There are various different tasks in the domain of food recommendation. *Image classification* tries to analyse a dish or single ingredient and connect that with additional information to compute a meaningful recommendation [54]. *Restaurant recommendations* are focused on giving the user a fitting recommendation on where to eat or order next, usually based on the restaurants' menus [32]. *Ingredient substitution* is interested in modifying dishes or recipes in a healthy manner [47]. *Recipe or dish recommendation* is focused on giving the user a whole meal plan or a single recipe or dish for self-preparation [10, 50]. There are of course various possible imaginable hybrid approaches like recommending a recipe based

on image classification of a picture of the users' fridge. In our work, we target mainly a healthy recipe recommendation task and we propose a connected data set that can be used for this purpose. However, as we provide meaningful mappings via connected external data sets (mainly the *FoodKG* [21]), the scope of the usability is larger.

Thus, our research question is *What data can support the ongoing research in the food recommendation area to improve comparability, realism, and health awareness?*. Initially, we see the greatest need for action on a common and, above all, public data basis. Related works [15, 26, 36] suffer from either a lack of features which could contribute to the *health awareness* aspect or a lack of meaningful user-item interactions. Also, most works do not provide linked data which can be used for generating recipe explanations in order to increase the users' trust. Moreover, the data sets/codes are often not publicly available or contain only small data. For instance, the recently shared *FoodKG* [21] knowledge graph does not contain any user-recipe interaction data.

In the scope of this paper, we refer to health awareness as not only recommending food items which minimize intake of calories but considering multiple nutrients and, in the best case, inferring a personalized nutritional need.

**Our contribution** is the publication of a linked, healthiness-aware, and argument-enabling recipe data set called *HUMMUS* (Health-aware User-centered recoMMendation and argUment enabling data Set). Moreover, we release the used crawler, related scripts for data cleaning & processing, the assessment of current related work, and a showcase of interesting use cases. The main features of the *HUMMUS* data set can be summarised as follows:

- *Linked*: We built our linked data set on top of a simplified version of the *FoodKG* [21] and additionally collected information, namely user and user-item interactions data from the *Food.com* website.
- *Healthiness-aware*: For each of the recipes, the healthiness scores are calculated according to the official guidelines from the WHO (the World Health Organisation) [22, 69] or the FSA (the UK Food Standards Agency) [16, 45] based on nutrition values. In addition, we implement a simplified version of the *NutriScore* [9].
- *Argument-enabling*: It is possible to reason over the data set in a semantic manner with e.g. SPARQL [1].
- *User-item interactions*: It is possible to use the data for collaborative recommendation as it introduces user-item interactions like recipe ratings, reviews, comments and authorships.

All resources are available through our GitLab repository<sup>1</sup>.

The remainder of the paper is structured as follows: In Section 2, we describe related fields with a focus on healthiness-aware, explainable recommendations, and food-related data sets and ontologies. Section 3 introduces in detail the creation and properties of our linked data set. Section 4 showcases some small-scale experiments and their evaluations on our data, while Section 5 concludes our paper and gives a brief outlook on future works.

<sup>1</sup><https://gitlab.com/felix134/connected-recipe-data-set>. The README.md file provides additional technical information regarding the data and steps to follow to reproduce the shown results, figures, and tables.

## 2 RELATED WORK

In this section, we give a brief overview of related works according to three axes: (1) food, recipe and nutrition data sets, (2) healthiness-aware food recommenders, and (3) explainable recommendation approaches. The axes are not mutually exclusive as some works tackle several aspects.

### 2.1 Food, Recipe & Nutrition Data Sets

Here, we overview frequently-used or otherwise interesting data sets and ontologies related to food. We analyse them with respect to the following requirements/desirable features: (1) being large-scale, (2) containing user-items interactions for user preference inference, (3) being connected/linked to enable the reasoning over it, (4) containing nutrition data allowing to evaluate the healthiness of a given recipe. Table 1 provides a basic comparison of the existing data sets that are detailed below.

*RecipeIM+* [31] is, to our best knowledge, the largest food-related data set in the field. It is composed of approximately 1 million recipes and 13.7 million connected images. The recipes originate from various different recipe websites. It lacks user-recipe interactions and semantic knowledge.

The *FoodOn* ontology [13] aims to categorize and group related food products in terms of their food type, food transformation process, consumer group, cultural origin, food packing and more. The most interesting relations for recipe recommendation are the product type relations. As an example, a *Fuji apple (whole)* is a subclass of *apple (whole)*, *apple (whole or parts)*, *apple food product*, *pomaceous fruit food product*, *plant fruit food product*, *plant food product*, and finally a subclass of *food product*. This ontology can be reasoned over using SPARQL [1, 2]. It consists of over 9 600 food products. The ontology alone of course does not include recipes or user-item interactions but could be exploited for additional knowledge.

The *FoodData* [51] (formerly known as *USDA* data set) provides an extensive collection of nutrient-related information (micro- and macro-nutrients like carbohydrates, vitamins, minerals and more) for single ingredients. It includes approximately 8 000 natural food products and 400 000 branded food products with each 46 measured nutrient features and more. It is important to note, especially for non-pre-processed nature food products, that the data only gives an average value. For products like an apple, this might lead to inaccuracies when used for the recommendation process, as nutrient values change during the cooking and preparation process, and products like apples vary in size. For industrial products like bagged soups, the problem is non-existent. Similar to the previously mentioned ontology, this data set is not directly usable for recommendations but might help as an additional data source.

The *FoodKG* [21] is a knowledge graph that provides connections between the *RecipeIM+* data set, the *FoodOn* ontology and a subset (*SR Legacy Foods*) of the *FoodData*. They achieved this goal by embedding ingredient names of all recipes and both to-be-connected data sets using natural language processing (NLP) techniques and then mapping similar entries. In order to enable the usage of semantic technologies, they transformed all their connected data into *RDF* triples [61]. We use the *FoodKG* as the base of our work as it provides rich food data. However, one has to keep in mind that

**Table 1: Food data set comparison**

Dataset	# items	# users	# user-item	linked data	nutrition data	comments
<i>Recipe1M+</i> [31]	1M recipes 13M images	N/A	N/A	N/A	FSA scores	Recipes paired with food images
<i>FoodOn</i> [13]	9.6K food items	N/A	N/A	yes	N/A	Food product categories as hierarchy
<i>FoodData</i> [51]	6K natural food products 400K branded products	N/A	N/A	N/A	avg, min, max amount nutrient per 100g	Food products with expanded nutrient profiles
<i>FoodKG</i> [21]	62M triples	N/A	N/A	yes	N/A	KG fusing <i>Recipe1M+</i> , <i>FoodOn</i> , and <i>FoodData</i>
Majumder et al. [30]	230K recipes	25K	1.1M reviews	N/A	N/A	Recipes and user reviews from <i>Food.com</i>
Trattner et al. [49]	~ 61K recipes	25K	17M bookmarks	N/A	WHO, FSA scores	Recipes and users from <i>All-recipes.com</i>
<i>MyFitnessPal</i> [58]	23.3K recipes	9.8K	156K diary records	N/A	nutrients	Recipes and users from <i>My-FitnessPal.com</i>

the linkage quality (with a coverage of 92% between recipe ingredients and *FoodOn* entries) of the *FoodKG* is not perfect and might be improved by more recent NLP approaches. Moreover, falsely connected entries might lead to inaccuracies when reasoning over the connected ontology. In total numbers, the data set consists of over 62 million triples. However, they are lacking user-recipe relations and in addition, did not import the ingredient’s nutrient information from *Recipe1M+*.

*RECIPTOR* [27] proposes a recipe representation learning model that uses triplet sample mining from *FoodKG*. It also provides the food tags from *Food.com* and its original recipe categorisation (9 categories in total).

*MyFitnessPal* [58] is a data set collected from *MyFitnessPal.com* (2014-2015) including 156 000 food diary entries from approximately 9 000 users and 23 300 recipes. It can be used to get user-item interactions. For instance, in [29] the authors use graph neural networks to learn joined recipe representations from *MyFitnessPal* and the *FoodKG*. They generate recipe recommendations and then try to find similar and healthy recipes which can be recommended instead.

Majumder et al. [30] aim at generating new recipes from incomplete user inputs. It is based on their collected data from *Food.com*, which includes approximately 230 000 recipes and 1.1 million reviews (2000-2018) serving as user-item interactions. However, it is not directly connected to semantic data and does not offer the original recipes’ URLs to connect them with a subset of the *FoodKG*, nor nutrients data.

Additional commonly used, but smaller data sets can be found in the work by Trattner et al. [49], crawled from *Allrecipes.com* (2000-2015). As a user-item interaction, the authors consider the action of bookmarking a recipe.

As it can be seen, there is still room for improvement in terms of available data that satisfies our aforementioned requirements. Our HUMMUS data set bridges this gap.

## 2.2 Health-Aware Food Recommendation

The works on (healthy) food recommendation (e.g. [18, 24, 36, 44]) mainly focus on recommending personalised recipes or diets. As argued in [10], most existing food recommenders ignore crucial health factors (e.g., allergies and nutrition needs), and/or do not consider the rich food knowledge for recommending healthy recipes. But as the awareness of food-related health and sustainability issues rises, especially bad dietary recommendations can result not only in a loss of trust towards the system but can also harm the user triggering physical and/or mental illness [67]. In this subsection, we discuss existing healthy food recommendation systems from the following perspectives: (1) their definition of healthiness, (2) their way of considering the healthiness dimension in the recommendation process, and (3) their response to some of the domain challenges.

**What is a healthy recipe?** The notion of *healthy food* has been evolving in time. Some works focus on calorie intake (e.g. [19, 43]), others on a cholesterol factor (e.g. [48]), or multi-criteria based on nutrients like protein, sodium, cholesterol, and saturated fats (e.g. [66]). Wang et al. [54] assign one or many nutritional benefit categories (weight loss, health care, nutritional supplements, and disease recovery) to each recipe. Moreover, they classify the recipes as suitable or unsuitable for a particular kind of user health tag (96 in total) based on their ingredients according to the acquired healthy diet tips (the source of which is not precise). Nowadays, we can also refer to the guidelines of national and international health organisations, like the World Health Organization (WHO) [39] or the French National Agency of Public Health [9, 11] or the UK Food Standards Agency (FSA) [3, 16, 45]. The idea of using food traffic light systems originated from such guidelines and calculating a single value based on them can be found in several works, e.g. [15, 49, 50]. However, such scores are not personalized and thus only express a general classification. In contrast to that, Chen et

al. [10] introduce nutrient and macronutrient budgets based on the guidelines from the American Diabetes Association (ADA) [5] that can be adjusted w.r.t. user's health condition. Moreover, some efforts have been done to target a specific audience based on their health conditions, like thyroid, diabetes, or high blood pressure patients [12, 37, 52]. However, as we are interested in a broader audience, it is not that easy to define personal medical requirements.

**How to take the healthiness dimension into account?** Hybrid approaches have shown their effectiveness in recommenders [26]. Those systems are a mixture of approaches like content-based filtering (recommendation only over item information and thus similar items), collaborative filtering (recommendation based on user-item interactions and thus similar users), knowledge-based systems (information retrieval via semantic data), and demographic-based systems (similar users based on demographic information). Most of the novel recommendation models belong to this group. The existing works differ in the way they incorporate the healthiness dimension in line with their definition of the latter. Thus, to consider numerical constraints w.r.t. the amount of a given nutrient, symbolic number comparison is applied in [10] resulting in a new sub-graph fed to a neural network-based system. To incorporate negative constraints, the authors use an embedding-based method for query expansion. To deal with nutrient-based constraints, Toledo et al. [66] use multi-criteria decision analysis. Reinforcement learning-based models can be used for single values. Thus, in [43] a reward is based on the recipes' calories. Wang et al. [54] propose an attention-based model. A post-processing procedure can also be applied to re-rank the results of the prediction based on their health scores (e.g. [15]).

**What are the challenges and the response to them?** Based on the state-of-the-art, we identify the following main challenges in the field of healthy food recommendation:

*Data availability & quality:* The lack of public data and/or poor data quality is one of the biggest issues of healthy food recommendation [36, 67]. Moreover, the data sets are non-standardised [26]. To collect more user context data, Min et al. [36] suggest using smartwatches to track the users' physical status or food histories to track eating habits. Chen et al. [10] randomly generate user dietary preferences and allergies. Wang et al. [54] assign health tags to users based on their personal information and their tweets. To simulate user-item interactions, they assign similar "health-tagged" recipes and users. The missing food preferences could be considered as one of the shortcomings of their solution.

*Cold-start problem:* Like in other recommendation domains, healthy food recommendation suffers from the cold-start problem [15, 43, 49, 50]. To overcome this issue, Rabbi et al. [43] suggest collecting food diary data for at least 7 days before actually starting the personalised recommendation process. Chen et al. [10] propose a knowledge-based question answering (KBQA) recommender allowing to consider explicit requirements from users, for example not only recommending a breakfast dish but one which includes bread, which helps to avoid the cold-start issue.

*Evaluation:* Most of the existing approaches are evaluated w.r.t. the accuracy of recommendations (e.g. [10, 54]). To account for the healthiness of the returned results, some healthiness-oriented metrics have been proposed like health-revised recall at top- $k$  [28]. Additionally, human evaluation can be performed and is a rather

desirable procedure (e.g. [10]), as well as case studies (e.g. [66]). Trattner et al. [15, 49, 50] point out that beyond accuracy goals (e.g. serendipity, diversity, etc.) are lacking in the food recommendation domain.

*Eating behaviour change:* Healthy food promotion is not limited to a recommendation of healthy recipes. It deals with behaviour change. To assist such a change, recommender systems can be used, providing a user with personalised content. However, to achieve the final goal of behaviour change, such a system should go far beyond a traditional recommender system. Little effort has been done so far in this direction (e.g. [14, 33, 40]), which is partially due to the need for long-term user studies in order to evaluate the effectiveness of a proposed approach.

### 2.3 Explainable recommendation

In this subsection, we briefly discuss the related work focusing on the explainability of recommender systems, not necessarily from the food domain, but keeping in mind our healthy food recommendation scenario. We argue that explainability is especially important in the domain of health-aware food recommenders as recommendations alone might not help the user to change his/her eating habits. Explaining recommendations might help to increase the transparency of the system and thus increase the users' trust. It is worth noting that other techniques such as persuasion and behaviour change methods can be used to further increase the effectiveness of such a recommender. However, those are out of the scope of this work.

Usually, two kinds of explanation approaches are distinguished. Post-hoc explanations [4, 7, 8, 41, 55, 63, 65] mean that explanations are generated after recommendations. On contrary, embedded explanations [53, 56, 57, 62, 64] are automatically inferred by the recommendation model. The main shortcoming of post-hoc explanations is often considered to be their inability to explain the models [68]. However, we argue that in the healthy-aware food domain, the explanation of the model itself is less important than the one of the recommendation results in a reasonable and actionable way - even if the explanation does not fit the model. Another drawback of post-hoc explanations is that they usually require an additional data structure for the reasoning, which in the case of our *HUMMUS* data set is the *FoodOn* ontology. The advantages of post-hoc explanations are the possibility to control those independently from the recommendation process, and thus arguably a potential quality.

As an example, a user study [25] showed, that a collaborative-filtering approach for recommendations performs usually best (discarding hybrid recommenders) in terms of accuracy. However, when users were asked about two types of recommendations (embedded and post-hoc) they preferred the latter. Their work shows, that an end user is not interested in user-centric or socio-centric aspects of the explanation, but more in item-centric ones.

To foster transparency and trust, Padhiar et al. [42] extend the *FoodKG* and constructs a *Food Explanation Ontology* in order to infer logical explanations from a recommendation. They use a post-hoc explanation design.

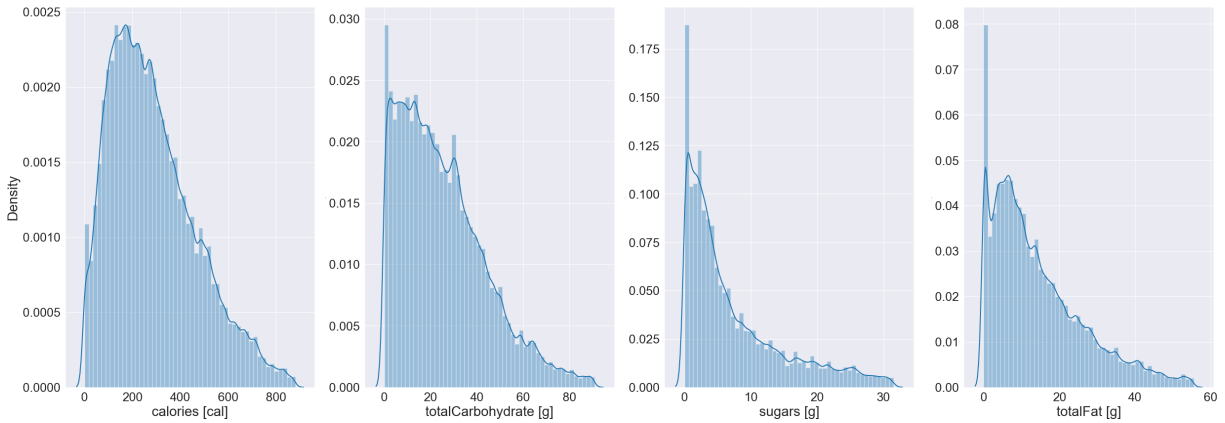


Figure 1: Density of nutrient values in the data set after normalization

### 3 THE HUMMUS DATA SET

As shown in the previous section, there is a need for a data set which can be used for an explainable and health-aware food recommendation. In order to support hybrid recommenders, we need both content-based features of the recipes, but also collaborative user-item interactions, which cannot be found to that extent in related work. Moreover, to enable post-hoc explanations, semantic data which is connected to the data set is provided.

In this paper, we present our *HUMMUS* (Health-aware User-centered recoMMendation and argUment enabling data Set) data set. It is primarily inspired by *FoodKG*. Our methodology consists of the following three steps. First, to ensure the *HUMMUS* connectivity with the *FoodKG*, we collect all unique recipe URLs from the *FoodKG*. For each of those recipes, we crawl additional data. Second, we engineer additional features like nutrition scores. Third, the crawled raw data is cleaned and pre-processed. This also includes the simplification of the *FoodKG* for further crossing and the integration of our data with the *FoodKG*. In the following, we describe in detail the collection process, the construction of the whole data set, and give an overview of the final data set.

#### 3.1 Data Collection

The *FoodKG* consists of approximately 1 million recipes from 21 different websites. Half of those recipes are from *Food.com*. Each other source from the *FoodKG* does not exceed 75 000 recipes. *Food.com* is one of the biggest recipe sites in English on the web. It currently consists of over 500 000 recipes including information like nutrients, preparation steps, ingredients, food-related tags, and metadata like creation date or average rating. There are currently over 300 000 public user profiles and over 2 million user-item interactions including ratings (with values from 1 to 5), comments, and authorship.

We collected additional information from all recipe URLs (approximately 500 000) from the *FoodKG* which are located on the domain *Food.com*. Thus, *HUMMUS* contains a larger spectre of data than the original *FoodKG*, and most importantly, it includes user-recipe interactions. We also explore new recipes not located in the *FoodKG* from each URL to a certain degree. In the end, we result in 507 335 recipes, 302 412 users, and 1 916 424 user-item interactions.

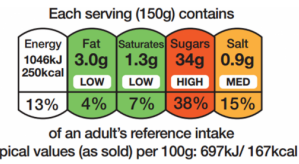


Figure 2: Exemplary traffic light labelling from the FSA<sup>2</sup>

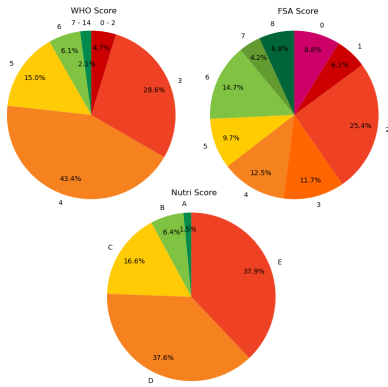
#### 3.2 Measuring Recipe Healthiness

Figure 1 shows the density of different nutrient values in the data set. All nutrient values are given in relation to one serving (even if the recipe might result in multiple servings). As an example, the WHO [59] recommends no more than 50g of sugar per day for adults. Additional health benefits can be inferred by limiting the daily intake to at most 25g of sugar per day. Depending on the number of meals per day, some recipes are clearly not healthy.

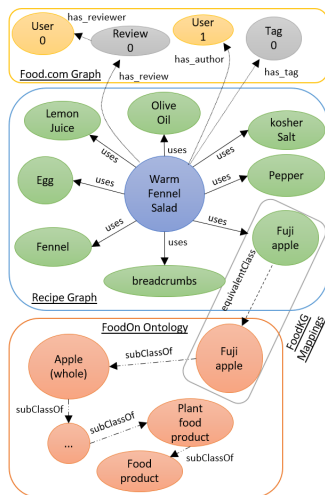
The question "How healthy is a recipe?" depends mainly on the composition of nutrients and personal medical requirements, and thus is rather complex. In order to measure the healthiness of the recipes, we deploy a system consisting of multiple *non-personalized scores*. These scores express the coarse healthiness which is focused on an average person. We do not provide *personalized scores*, which are more accurate for a given user, as we do not possess medical information about the users in the data set.

Similar to Trattner et al. [15, 49, 50], we take the food traffic light systems from national and international health agencies like the FSA [16, 45], WHO [22, 69] and use those to compute a single value expressing the overall *healthiness of a recipe*. Those traffic light systems are usually used for front package labelling and can look like shown in Figure 2. They offer a reference for the most nutrients and the energy value encoded as colours ranging from green to red. All colours (depending on the model 3 or 5) get encoded in numbers. The final score is the sum over all categories of the encoded colours. To improve the precision of those scores, we extend them by linear scaling and normalization. The scaling allows a higher precision, as we do not score a fixed value for each category, but instead scale

<sup>2</sup>Retrieved from <https://www.food.gov.uk/safety-hygiene/check-the-label>



**Figure 3: The WHO score, FSA score, and NutriScore rating the healthiness of each recipe in the HUMMUS data set. For the NutriScore, the colour encoding follows its specifications. For the FSA score, the colour encoding is a similar split between the colour shades red, orange, and green. The WHO score is not evenly coloured but follows the colours of the distributions of the NutriScore.**



**Figure 4: Exemplary structure of the simplified FoodKG. Structure-wise irrelevant relations are discarded.**

it dependent on the distance to each bound. The normalization has the effect that the resulting score is in the closed interval of 0 and 1.

In addition to those two scores, we add the simplified version of the NutriScore [9]. An advantage of this score is that it does not only express the healthiness of multiple nutrients but already compresses this information into one single score. This score was published by medical researchers and has a more complicated method to add all single-nutrient-scores together, thus the resulting value might be more accurate than for both previous scores. On the other side, we do only offer the simplified score, as the original NutriScore takes additional information into account. For instance, it is differently defined for recipes representing dishes or beverages and has special handling for cheese, fruit and vegetable foods. We plan on adding

the additional calculations in future, as the semantic structure of our data set allows the querying for those required features.

In order to get an overview of the healthiness of our data set, we plot the NutriScore, WHO score, and FSA score for each recipe in the data set, as seen in Figure 3. The scores cannot be compared directly because of the different intaken arguments and resulting categories. However, all three scores express a similar tendency - most recipes are not healthy at all. Depending on which categories we label as 'healthy', the number of healthy recipes in the data set changes. The WHO Score rates the recipes in categories from 0 to 14, while a higher number means healthier. If we assume that the scores from 0 to 4 are not healthy and remove them, only 23.2% of the total recipes remain. The FSA Score rates the recipes from 0 to 8, while a higher number also means healthier. With the assumption of unhealthy recipes from 0 to 4, we result in 35.4% neutral or healthy recipes. The NutriScore defines the categories D and E as unhealthy, thus we do not need to make any assumption. With that, 24, 5% of all recipes are neutral or healthy. The scores shown in Figure 3 are encoded according to those mentioned assumptions.

### 3.3 FoodKG Simplification and Extension

Next, we simplify the FoodKG. Thus, we discard all uninteresting triples for our recommendation scenario. As described in Section 2.1, the FoodKG is built upon Recipe1M+, FoodData, and FoodOn. As the total recipe nutrient value of recipes cannot be calculated as a simple sum of its ingredients, we discard the connections to the ingredient-based nutrient set FoodData. In HUMMUS, the only source of the recipes is Food.com. We map their ingredients nodes directly to FoodOn food categories. So, we can remove all provenance nodes present in the original FoodKG. Next, we add all the additional features we got from our crawler to the graph, as well as the nutrition scores calculated on the previous step. Finally, we remove all nodes which are not connected to our data set.

```

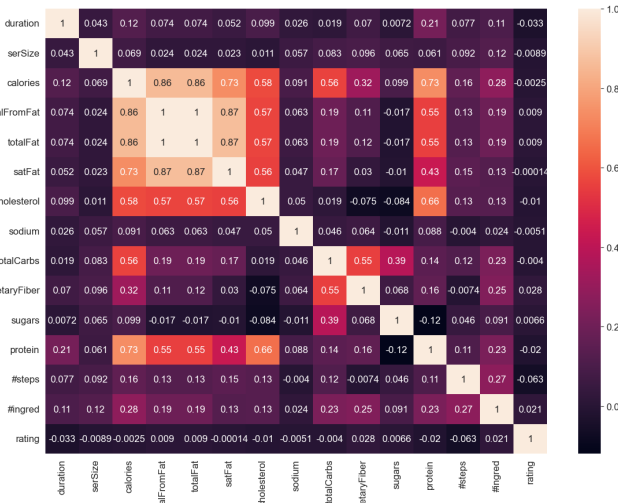
PREFIX foodkg: <http://idea.rpi.edu/heals/kb/>
PREFIX foodon: <http://purl.obolibrary.org/obo/FOODON_>
SELECT DISTINCT ?title ?recipe
WHERE {?recipe a foodkg:recipe;
      rdfs:label ?title.
      FILTER NOT EXISTS { #invertebrate animal food product
        ?recipe foodkg:uses/owl:equivalentClass/rdfs:subClassOf* foodon:00001176.}
      FILTER NOT EXISTS { #vertebrate animal food product
        ?recipe foodkg:uses/owl:equivalentClass/rdfs:subClassOf* foodon:00001092.}
      FILTER NOT EXISTS { #seafood product
        ?recipe foodkg:uses/owl:equivalentClass/rdfs:subClassOf* foodon:00001046.}
      FILTER NOT EXISTS { #animal
        ?recipe foodkg:uses/owl:equivalentClass/rdfs:subClassOf* foodon:00003004.}
} LIMIT 100
    
```

**Listing 1: SPARQL query to extract 100 vegetarian recipes from the simplified FoodKG. Omitted standard prefixes rdfs and owl.**

An example of the final simplified FoodKG can be seen in Figure 4. Note, that all relations which are not essential to the graph's structure are not shown. Those include the same information as provided by the data set like nutrition for recipes or the review's rating and text. In this example, we only show the connections

**Table 2: Data set size and sparsity after *threshold discarding* with different parameters:  $k = (k_{users}, k_{recipes})$** 

$k$	#recipes	#users	#interac.	sparsity [%]
(1, 1)	507 335	302 412	1 916 424	99.99875
(2, 2)	270 284	83 881	1 476 615	99.99348
(10, 10)	30 859	18 316	601 887	99.89351
(50, 2)	254 741	4 614	1 015 399	99.91361
(100, 2)	245 499	2 376	880 154	99.84910
(2, 50)	2 987	52 509	294 058	99.81251
(2, 100)	1 091	42 968	187 327	99.60039

**Figure 5: Heatmap of a subset of features of the data set. The data behaves as expected: Carbohydrates, fats and calories are correlated.**

between one ingredient (Fuji apple) and the corresponding *FoodOn* node. However, most ingredients have such a connection, which makes it possible to reason over the graph. Similar holds for users and recipes. Users might have multiple connected reviews. Recipes are possibly connected to multiple reviews, tags, and one author. The resulting knowledge graph can be queried using SPARQL. For instance, a query as seen in Listing 1 searches for all vegetarian recipes (recipes whose ingredients are only of class *Vegetarian food product* or its sub-classes).

### 3.4 Final Connected Data Set

Our published data set *HUMMUS* is similar to the one released by Majumder et al. [30], but improves the data quality on multiple occasions. Both are solely collected from *Food.com*, but ours is more recent and thus provides more data. Moreover, it includes more features for recipes, reviews and users. Most importantly, we provide a direct mapping to a subgraph of the *FoodKG* and thus enable semantic reasoning, especially over the *FoodOn* ontology.

Depending on the use case of the data, there might be a required density of the data. A high level of sparsity means that for an average user or recipe, we have a small number of interactions.

To overcome this issue, one can discard recipes or users with a small count of interactions. However, the more users or items are discarded, the more unrealistic the experiments get. In a realistic scenario, we also need to deal with the cold start problem; namely how new users (or those with a small number of interactions) can also be served with good recommendations. The usual way is to use content-based recommendations for those. We implement a *threshold discarding* technique allowing us to keep the data satisfying the threshold conditions only and thus reducing the data set sparsity. *Threshold discarding* with parameters  $k_{recipes} = 2$  and  $k_{users} = 2$  means that we get rid of every user (respectively recipe) which has less than two interactions. Depending on the required setup, different parameters can be chosen. Table 2 shows the results on the data set when modifying the *threshold discarding* parameters. It is noticeable that a small proportion of users heavily interacted with items, while the majority only interacted with one or two recipes. Depending on the *threshold discarding* parameters, the composition of the number of unique recipes and users present in the interactions changes. Thus, the pruning, depending on the use case, might change the rankings of recommended recipes [6]. As a result, we grant access to the raw data, a version of the preprocessed data set, and the data pruning code.

We explore the correlations between a subset of *HUMMUS* features. To get an idea of correlated features refer to Figure 5. It shows a heatmap composed of the nutrient-based features, the number of ingredients and instruction steps, cooking duration, and the rating. The correlations are straightforward, as strong ones, for example, can be found between *calories* and *protein*, *saturated fat*, and *total carbohydrates*. Also, *total fat*, *proteins*, *cholesterol*, and *saturated fat* are correlated with each other. Weaker correlations can be found between *sugar* and *total carbohydrates*, and between the *number of ingredients* and all nutrient-based features. Protein-containing recipes in our data set seem to be weakly correlated with the cooking duration. The user ratings are, as expected, not correlated at all with any other feature.

In comparison to other well-known data sets in the recommendation area, the sparsity is a lot higher. *MovieLens100k* [20] for example has a sparsity of 93, 69%. This issue leads to poor performance of recommendation methods which focus only on user-item interactions, demonstrated in the next section.

To sum up, the *HUMMUS* data set features realistic recipe data most importantly including user-item interactions and nutrition information. The set is relatively sparse in comparison to other sets used in the recommendation area, which indicates the challenge when using collaborative recommenders. Lastly, we simplified and extended the *FoodKG* with our data, in order to support semantic reasoning for multiple tasks like explanations or search queries.

## 4 EXPERIMENTS

This section describes a set of experiments conducted on the user-recipe interactions of our *HUMMUS* data set only. The goal we set ourselves was rather to highlight the possibilities and challenges the data set offers, than scoring the best possible results. Thus, we do not perform a detailed hyper-parameter tuning for only a few methods and instead perform the experiments on multiple different models with standard hyper-parameters. We chose standard models like



**Table 3: Resulting scores for our data set with a sparsity of 99.89351% of the processed models from different python libraries (*implicit* [17], *irec* [23], *recommenders* [34]).**

<i>Model</i>	<i>Library</i>	<b>MAP</b>	<b>nDCG@10</b>	<b>Precision@10</b>	<b>Recall@10</b>
<b>Random</b>	<i>irec</i>	N/A	N/A	0.000410	0.000300
<b>ICTR</b>	<i>irec</i>	N/A	N/A	0.000546	0.000491
<b><math>\epsilon</math>-Greedy</b>	<i>irec</i>	N/A	N/A	0.001392	0.001149
<b>kNNBandit</b>	<i>irec</i>	N/A	N/A	0.005051	0.003750
<b>ThompsonSampling</b>	<i>irec</i>	N/A	N/A	0.009746	0.006947
<b>PTS</b>	<i>irec</i>	N/A	N/A	0.010483	0.008021
<b>LogPopEnt</b>	<i>irec</i>	N/A	N/A	0.014305	0.012201
<b>Entropy</b>	<i>irec</i>	N/A	N/A	0.025334	0.020367
<b>BestRated</b>	<i>irec</i>	N/A	N/A	0.025334	0.020367
<b>MostPopular</b>	<i>irec</i>	N/A	N/A	0.025471	0.020468
<b>ALS</b>	<i>implicit</i>	0.010786	0.022453	0.031556	N/A
<b>ALS_nmslib</b>	<i>implicit</i>	0.009061	0.018922	0.027670	N/A
<b>ALS_annoy</b>	<i>implicit</i>	0.009380	0.019188	0.026773	N/A
<b>Tf-idf</b>	<i>implicit</i>	0.005370	0.009654	0.012213	N/A
<b>Cosine</b>	<i>implicit</i>	0.004210	0.007412	0.009193	N/A
<b>LFM</b>	<i>implicit</i>	0.003492	0.007988	0.011675	N/A
<b>BM25</b>	<i>implicit</i>	0.004422	0.007593	0.009672	N/A
<b>Item-item</b>	<i>implicit</i>	0.011759	0.023017	0.030674	N/A
<b>BPR</b>	<i>recommenders</i>	<b>0.023879</b>	<b>0.062695</b>	<b>0.043514</b>	<b>0.058164</b>
<b>BiVAE</b>	<i>recommenders</i>	0.020657	0.057688	0.042088	0.051492

*BPR*, *ALS*, models used in the related work like the reinforcement learning model *kNNBandit*, and simple content-based models like *Tf-idf* for comparability. Also, we show that the average nutrition score is very low or unhealthy out-of-the-box.

#### 4.1 Setting

The setting of all experiments is the same. First, we take a subset of the data utilizing *threshold discarding* with the parameters  $k_{users} = 10$   $k_{recipes} = 10$ . Second, in order to compare all results, we *only* take the user-recipe interactions (a set of ratings, each described by a value  $v$  from a user  $u$  for a recipe  $r$ ) as features into consideration. Last, for each method, we use a 75%/25% train/test split and the top 10 recommendations only. For the healthiness evaluation, we choose the average *NutriScore* of the top 10 recommendations. As a baseline, we chose a random top 10 list of recipe recommendations. We use a set of python libraries offering the models' implementations, namely *implicit* [17], *irec* [23], and *recommenders* [34]. All models are processed with their standard or typical hyperparameters.

In order to promote healthy recipes more prominently, we implemented three post-processing methods, namely *filter by threshold* (removes recipes below a threshold), *filter by percentage* (removes the  $x\%$  unhealthiest recipes), and *replace unhealthy recipes* (finds similar but healthier recipes based on cosine similarity using a Tf-Idf embedding). All methods modify the list of recommendations of a given user and a given nutrition score.

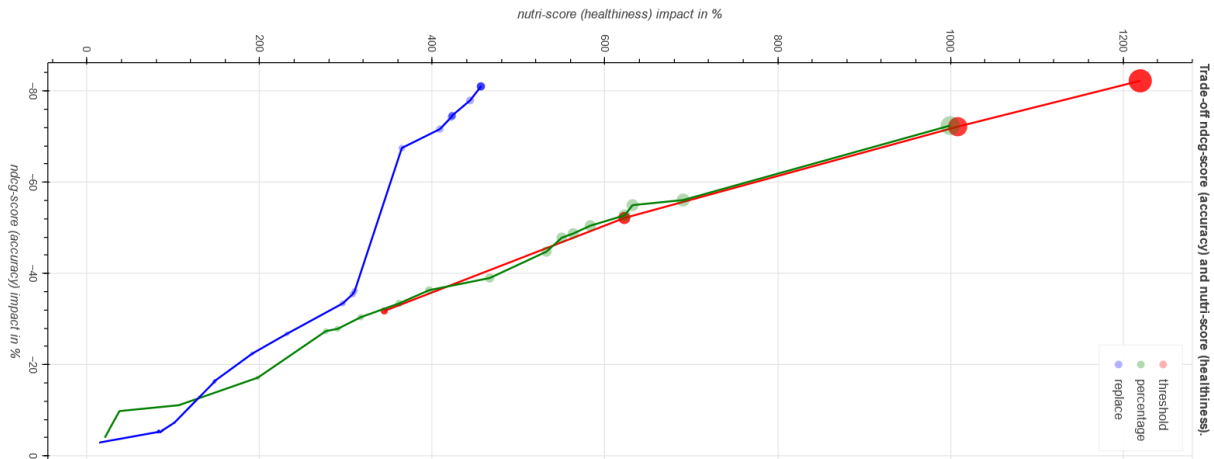
#### 4.2 Results

For the results without post-processing refer to Table 3. The *Random* baseline performs as expected the worst with very low precision and recall scores. *ICTR*, a bandit-based regression model, did only result slightly better, and thus cannot be considered as a reasonable

**Table 4: The impact of each post-processing method, namely *filter by threshold* with threshold (T), *filter by percentage* with percentage (P), and *replace unhealthy recipes* with cosine similarity threshold (CS), in % on the accuracy (nDCG@10) and the healthiness (average *NutriScore*) on the baseline with the BPR model. There are only sample parameters given to show the trade-off at around 30% reduction of accuracy.**

Impact on baseline in %	nDCG@10	Avg FSA
BPR	0.062695	2.297
T=0.1	-31.77%	+344.45%
P=0.6	-30.40%	+317.14%
CS=0.13	-33.42%	+296.27%

recommender for our scope. Next, in ascending order regarding the precision score, the regression-based models [46] using methods like  *$\epsilon$ -Greedy*, *kNNBandit*, *ThompsonSampling*, *PTS*, and *LogPopEnt* perform way better as the baseline, with a precision of approximately 1%. Those are comparable, in terms of precision, to the nearest neighbours models like *Tf-idf*, *Cosine*, and *BM25*. Also, *Logistic Matrix Factorization (LFM)* yields similar results. Models being in the precision range around 2% and 3% are the regression models based on *Entropy*, *BestRated*, *MostPopular*, the *Alternating Least Squares (ALS)* model and its approximations *ALS\_nmslib* and *ALS\_annoy*, and the nearest neighbour model *Item-Item* recommender. The best-performing models by far are the *Bayesian Personalized Ranking (BPR)* and the *Bilateral Variational Autoencoder (BiVAE)* models with scores around 5% respectively 4%. As already mentioned in the previous section, all results are not fine-tuned and thus only give an approximate overview of the models' performances.



**Figure 6: Trade-off between the accuracy (nDCG@10) and the *NutriScore* for each post-processing method compared to the BPR baseline. The size of the data points reflects the combined trade-off impact (accuracy impact plus healthiness impact).**

The results after post-processing, using BPR as the baseline mode, are shown in Table 4 and Figure 6. The table shows exemplary results focused on the *FSA score* for one parameter configuration, whereas the picture is focused on the *NutriScore* for multiple configurations. The trade-off between accuracy and healthiness is present in all results, no matter which score is focused. In general, the more the healthiness of the recommendations is promoted, the lower the accuracy is. However, both filtering methods, threshold and percentage, perform very similarly and are better than the replacement method at an accuracy impact of around 15%. Before this value, the replacement method delivers faster growth in terms of improved nutrition scores.

### 4.3 Discussion

Derived from the results it can be certainly concluded that a simple recommender, based only on user-item interactions, is not sufficient in the scope of the presented data set and the healthy food recommendation setting.

The trade-off between accuracy and healthiness was expected. However, it is unclear how to conduct the right balance as it might depend on each user’s preferences. For small changes in regard to the accuracy, the replacement method should be utilized, and vice-versa for bigger ones the threshold or percentage filtering method. Moreover, the impact of the methods’ parameters varies depending on the targeted nutrition score. Especially the *NutriScore* with a small average score on the baseline seems to result in *better* trade-offs.

**What can be done to improve the recommender?** First of all, in order to solve the cold-start problem for potential new users and also improve the recommenders’ performance, a hybrid recommender should be deployed. Thus, additional information and features should be taken into consideration. Those features could include the recipes’ tags, the food categories derived from the *FoodOn* ontology, a review text embedding, the steps to take for cooking the recipe, and more. This motivates further the additional features and connected properties of our data set.

Secondly, it might be useful to take, on top of *non-personalized scores*, also *personalized scores* into consideration. However, they can only be used if medical properties like the Body Mass Index (BMI), height, weight, age, and sex of individuals are known. One of those scores is published by Miffil et al. [35] and takes all mentioned features in addition to an activity factor and smoke factor into account.

Third and lastly, the connected semantic data from the *FoodKG* can be exploited to enable meaningful explanations.

## 5 CONCLUSION AND FUTURE WORK

Obesity and overweight are enormous problems worldwide. To address the sub-problem of healthy nutrition, we set our focus on the area of food recommendation with the research question *What data can support the ongoing research in the food recommendation area to improve comparability, realism, and health awareness?* The problem we address is, that high-quality data sets in the domain of healthy food recommendations are either not publicly available, lack certain kinds of features, or are of a too-small size. We propose *HUMMUS*, a connected data set collected from *Food.com*, including rich information for recipes, users, and reviews. Those features include standard recipe information, nutrient information, ratings and text reviews, user profile descriptions and more. The contributed data is connected to the *FoodKG* [21] including the *FoodOn* ontology which can be used to reason over various food categories describing ingredients of our data set. Moreover, we provided an in-depth analysis of our data set, which has led us to believe, that one key challenge in the healthy food recommendation domain is to work with data sets consisting mainly of unhealthy recipes. In addition, we extended the idea of *nutrition scores* and generated those as features for each recipe, expressing a general healthiness of dishes. However, for the personalized nutrition scores, we lack medical data and thus needed to generate user features like the *BMI*.

We conclude, that, although our experiments did not perform as well as hoped, they show the necessity of taking additional features into consideration when dealing with that level of sparse

data. However, as our data set is so rich in multiple features, we did not explore all possibilities the data set offers. Thus, we would like to open this challenge by publishing the data with free access.

For future works, there are multiple directions which are imaginable with our data set. First, a *Healthiness-aware recipe recommendation* might be realised with a hybrid recommender including multiple features and a re-ranking strategy based on personalized or un-personalized nutrition scores. Second, path-based recommendations might also be worth a try exploring. The paths could be a combination of the user-recipe interactions and the connected *FoodOn* ontology expressing an ingredient type hierarchy. A possible advantage of those methods might be the implicit processing of explanations which might foster the trust of a user in the recommender. Third, *Post-hoc Explanations* while reasoning over the same Knowledge Graph might be interesting, especially in comparison to the path-based approaches. Next, exploring text-based features like the users' profile descriptions, recipe descriptions and directions, or text reviews might be interesting when working in the area of NLP. Last, including *Image Classification* in a possible recommender, might also be possible, as the data set is connected to the *FoodKG* which again is connected to the *Recipe1M+* data set.

## REFERENCES

- [1] 2013. *SPARQL 1.1 Overview*. W3C Recommendation. W3C. <https://www.w3.org/TR/2013/REC-sparql11-overview-20130321/>.
- [2] 2021. *RDF-star and SPARQL-star*. Final W3C Community Group Report. W3C Community Group. <https://w3c.github.io/rdf-star/cg-spec/2021-12-17.html>.
- [3] Food Standards Agency. 2018. Nutrition labelling. <https://www.food.gov.uk/business-guidance/nutrition-labelling>
- [4] Qingyao Ai, Wahid Azizi, Xu Chen, and Yongfeng Zhang. 2018. Learning Heterogeneous Knowledge Base Embeddings for Explainable Recommendation. *Algorithms* 11, 9 (sep 2018), 137. <https://doi.org/10.3390/a11090137> arXiv:1805.03352
- [5] American Diabetes Association. 2019. 5. Lifestyle Management: *Standards of Medical Care in Diabetes—2019*. *Diabetes Care* 42, Supplement\_1 (Jan. 2019), S46–S60. <https://doi.org/10.2337/dc19-S005>
- [6] Joeran Beel and Victor Brunel. 2019. Data pruning in recommender systems research: Best-practice or malpractice? *CEUR Workshop Proceedings* 2431 (2019), 26–30.
- [7] Bhavya and ChengXiang Zhai. 2020. Explanation Mining. In *Proceedings of the Seventh ACM Conference on Learning @ Scale*. ACM, New York, NY, USA, 321–324. <https://doi.org/10.1145/3386527.3406738>
- [8] Rose Catherine, Kathryn Mazaitis, Maxine Eskenazi, and William Cohen. 2017. Explainable Entity-based Recommendations with Knowledge Graphs. *CEUR Workshop Proceedings* 1905, Rule 2 (jul 2017), 2–3. arXiv:1707.05254 <http://arxiv.org/abs/1707.05254>
- [9] Julia Chantal and Serge Hercberg. 2017. Development of a new front-of-pack nutrition label in France: the five-colour Nutri-Score. *Public health panorama* 03, 04 (2017), 712–725.
- [10] Yu Chen, Ananya Subburathinam, Ching Hua Chen, and Mohammed J. Zaki. 2021. Personalized Food Recommendation as Constrained Question Answering over a Large-scale Food Knowledge Graph. *WSDM 2021 - Proceedings of the 14th ACM International Conference on Web Search and Data Mining* (2021), 544–552. <https://doi.org/10.1145/3437963.3441816> arXiv:2101.01775
- [11] Agence Nationale de Santé Publique. 2023. Nutri-Score. <https://www.santepubliquefrance.fr/determinants-de-sante/nutrition-et-activite-physique/articles/nutri-score>
- [12] Vaishali M. Deshmukh and Samiksha Shukla. 2021. Content-Restricted Boltzmann Machines for Diet Recommendation. In *Data Science and Security*, Samiksha Shukla, Aynur Unal, Joseph Varghese Kureethara, Durgesh Kumar Mishra, and Dong Seog Han (Eds.). Vol. 290. Springer Singapore, Singapore, 114–121. [https://doi.org/10.1007/978-981-16-4486-3\\_12](https://doi.org/10.1007/978-981-16-4486-3_12)
- [13] Damion M. Dooley, Emma J. Griffiths, Gurinder S. Gosal, Pier L. Buttigieg, Robert Hoehndorf, Matthew C. Lange, Lynn M. Schriml, Fiona S.L. Brinkman, and William W.L. Hsiao. 2018. Food on: A harmonized food ontology to increase global food traceability, quality control and data integration. *npj Science of Food* 2, 1 (2018), 1–10. <https://doi.org/10.1038/s41538-018-0032-6>
- [14] Ayoub El Majjodi, Alain D. Starke, and Christoph Trattner. 2022. Nudging Towards Health? Examining the Merits of Nutrition Labels and Personalization in a Recipe Recommender System. In *Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization*. ACM, Barcelona Spain, 48–56. <https://doi.org/10.1145/3503252.3531312>
- [15] David Elsweiler, Christoph Trattner, and Morgan Harvey. 2017. Exploiting food choice biases for healthier recipe recommendation. *SIGIR 2017 - Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval* (2017), 575–584. <https://doi.org/10.1145/3077136.3080826>
- [16] Food Standards Agency. 2016. Guide to creating a front of pack (FoP) nutrition label for pre-packed products sold through retail outlets. *Food Standards Agency November* (2016), 33. [https://www.food.gov.uk/sites/default/files/media/document/fop-guidance\\_0.pdf](https://www.food.gov.uk/sites/default/files/media/document/fop-guidance_0.pdf)
- [17] Ben Frederickson. 2023. *Implicit Recommender GitHub Page*. <https://github.com/benfred/implicit>
- [18] Xiaoyan Gao, Fuli Feng, Heyan Huang, Xian-Ling Mao, Tian Lan, and Zewen Chi. 2022. Food recommendation with graph convolutional network. *Information Sciences* 584 (Jan. 2022), 170–183. <https://doi.org/10.1016/j.ins.2021.10.040>
- [19] Mouzhi Ge, Francesco Ricci, and David Massimo. 2015. Health-aware Food Recommender System. In *Proceedings of the 9th ACM Conference on Recommender Systems*. ACM, Vienna Austria, 333–334. <https://doi.org/10.1145/2792838.2796554>
- [20] GroupLens. 1998. *MovieLens 100K Dataset*. <https://grouplens.org/datasets/movielens/100k/>
- [21] Steven Haussmann, Oshani Seneviratne, Yu Chen, Yarden Ne'eman, James Codella, Ching-Hua Chen, Deborah L. McGuinness, and Mohammed J. Zaki. 2019. FoodKG: A Semantics-Driven Knowledge Graph for Food Recommendation. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Vol. 11779 LNCS. 146–162. [https://doi.org/10.1007/978-3-030-30796-7\\_10](https://doi.org/10.1007/978-3-030-30796-7_10)
- [22] Simon Howard, Jean Adams, and Martin White. 2012. Nutritional content of supermarket ready meals and recipes by television chefs in the United Kingdom: cross sectional study. *BMJ* 345, dec14 14 (dec 2012), e7607–e7607. <https://doi.org/10.1136/bmj.e7607>
- [23] irec.org. 2023. *IRec GitHub Page*. <https://github.com/irec-org/irec>
- [24] Nan Jia, Jie Chen, and Rongzheng Wang. 2022. An attention-based convolutional neural network for recipe recommendation. *Expert Systems with Applications* 201 (Sept. 2022), 116979. <https://doi.org/10.1016/j.eswa.2022.116979>
- [25] Pigi Kouki, James Schaffer, Jay Pujara, John O'Donovan, and Lise Getoor. 2019. Personalized explanations for hybrid recommender systems. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*, Vol. Part F1476. ACM, New York, NY, USA, 379–390. <https://doi.org/10.1145/3301275.3302306>
- [26] Akshi Kumar, Pulkit Tanwar, and Saurabh Nigam. 2016. Survey and evaluation of food recommendation systems and techniques. *Proceedings of the 10th IN-DIACom; 2016 3rd International Conference on Computing for Sustainable Global Development, INDIACom 2016* (2016), 3592–3596.
- [27] Diya Li and Mohammed J. Zaki. 2020. RECIPTOR: An Effective Pretrained Model for Recipe Representation Learning. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, New York, NY, USA, 1719–1727. <https://doi.org/10.1145/3394486.3403223>
- [28] Diya Li, Mohammed J. Zaki, and Ching-hua Chen. 2023. Health-guided recipe recommendation over knowledge graphs. *Journal of Web Semantics* 75 (Jan. 2023), 100743. <https://doi.org/10.1016/j.websem.2022.100743>
- [29] Diya Li, Mohammed J. Zaki, and Ching hua Chen. 2023. Health-guided recipe recommendation over knowledge graphs. *Journal of Web Semantics* 75 (2023), 100743. <https://doi.org/10.1016/j.websem.2022.100743>
- [30] Bodhisattwa Prasad Majumder, Shuyang Li, Jianmo Ni, and Julian McAuley. 2019. Generating Personalized Recipes from Historical User Preferences. *EMNLP-IJCNLP 2019 - 2019 Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing, Proceedings of the Conference* (aug 2019), 5976–5982. <https://doi.org/10.18653/v1/d19-1613> arXiv:1909.00105
- [31] Javier Marin, Arif Biswas, Ferda Ofli, Nicholas Hynes, Amaia Salvador, Yusuf Aytar, Ingmar Weber, and Antonio Torralba. 2019. Recipe1M+: A Dataset for Learning Cross-Modal Embeddings for Cooking Recipes and Food Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 43, 1 (jan 2019), 187–203. <https://doi.org/10.1109/TPAMI.2019.2927476> arXiv:1810.06553
- [32] Jermaine Marshall, Priscilla Jimenez-Pazmino, Ronald Metoyer, and Nitesh V. Chawla. 2022. A Survey on Healthy Food Decision Influences Through Technological Innovations. *ACM Transactions on Computing for Healthcare* 3, 2 (2022). <https://doi.org/10.1145/3494580>
- [33] Jermaine Marshall, Priscilla Jimenez-Pazmino, Ronald Metoyer, and Nitesh V. Chawla. 2022. A Survey on Healthy Food Decision Influences Through Technological Innovations. *ACM Transactions on Computing for Healthcare* 3, 2 (April 2022), 1–27. <https://doi.org/10.1145/3494580>
- [34] Microsoft. 2023. *Microsofts recommenders GitHub Page*. <https://github.com/microsoft/recommenders>
- [35] D. Mifflin, Sandra A. Daugherty, Lisa A. Hill, Sachiko T. St. Jeor, and Barbara J. Scott. 1990. A new predictive equation in healthy individuals. *The American Journal of Clinical Nutrition* 51, 2 (1990), 241–241.

- [36] Weiqing Min, Shuqiang Jiang, and Ramesh Jain. 2020. Food Recommendation: Framework, Existing Solutions, and Challenges. *IEEE Transactions on Multimedia* 22, 10 (oct 2020), 2659–2671. <https://doi.org/10.1109/TMM.2019.2958761> arXiv:1905.06269
- [37] Divya Mogaveera, Vedant Mathur, and Sagar Waghela. 2021. E-Health Monitoring System with Diet and Fitness Recommendation using Machine Learning. *Proceedings of the 6th International Conference on Inventive Computation Technologies, ICICT 2021* (2021), 694–700. <https://doi.org/10.1109/ICICT50816.2021.9358605>
- [38] OECD. 2019. *The Heavy Burden of Obesity*. 256 pages. <https://doi.org/https://doi.org/10.1787/67450d67-en>
- [39] World Health Organization. 2020. Healthy diet. <https://www.who.int/news-room/fact-sheets/detail/healthy-diet>
- [40] Rita Orji, Lennart E. Nacke, and Chrysanne Di Marco. 2017. Towards Personality-driven Persuasive Health Games and Gamified Systems. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, Denver Colorado USA, 1015–1027. <https://doi.org/10.1145/3025453.3025577>
- [41] Ishita Padhiar, Oshani Seneviratne, Shruthi Chari, Dan Gruen, and Deborah L. McGuinness. 2021. Semantic Modeling for Food Recommendation Explanations. In *2021 IEEE 37th International Conference on Data Engineering Workshops (ICDEW)*. IEEE, 13–19. <https://doi.org/10.1109/ICDEW53142.2021.00010> arXiv:2105.01269
- [42] Ishita Padhiar, Oshani Seneviratne, Shruthi Chari, Dan Gruen, and Deborah L. McGuinness. 2021. Semantic Modeling for Food Recommendation Explanations. *Proceedings - 2021 IEEE 37th International Conference on Data Engineering Workshops, ICDEW 2021* (2021), 13–19. <https://doi.org/10.1109/ICDEW53142.2021.00010> arXiv:2105.01269
- [43] James M. Rehg, Susan A. Murphy, and Santosh Kumar. 2017. *Mobile Health*. Springer International Publishing, Cham. 1–542 pages. <https://doi.org/10.1007/978-3-319-51394-2>
- [44] Mehrdad Rostami, Mourad Oussalah, and Vahid Farrahi. 2022. A Novel Time-Aware Food Recommender-System Based on Deep Learning and Graph Clustering. *IEEE Access* 10 (2022), 52508–52524. <https://doi.org/10.1109/ACCESS.2022.3175317>
- [45] Gary Sacks, Mike Rayner, and Boyd Swinburn. 2009. Impact of front-of-pack 'traffic-light' nutrition labelling on consumer food purchases in the UK. *Health Promotion International* 24, 4 (dec 2009), 344–352. <https://doi.org/10.1093/heapro/dap032>
- [46] Javier Sanz-Cruzado, Pablo Castells, and Esther López. 2019. A simple multi-armed nearest-neighbor bandit for interactive recommendation. *RecSys 2019 - 13th ACM Conference on Recommender Systems* (2019), 358–362. <https://doi.org/10.1145/3298689.3347040>
- [47] Sola S. Shirai, Oshani Seneviratne, Minor E. Gordon, Ching Hua Chen, and Deborah L. McGuinness. 2020. Semantics-Driven ingredient substitution in the foodKG? *CEUR Workshop Proceedings 2721* (2020), 243–247.
- [48] Alain Starke, Christoph Trattner, Hedda Bakken, Martin Johannessen, and Vegard Solberg. 2021. The cholesterol factor: Balancing accuracy and health in recipe recommendation through a nutrient-specific metric. *CEUR Workshop Proceedings 2959* (2021), 1–11.
- [49] Christoph Trattner and David Elswiler. 2017. Food Recommender Systems: Important Contributions, Challenges and Future Research Directions. (2017), 1–16. arXiv:1711.02760 <http://arxiv.org/abs/1711.02760>
- [50] Christoph Trattner and David Elswiler. 2017. Investigating the Healthiness of Internet-Sourced Recipes. In *Proceedings of the 26th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, 489–498. <https://doi.org/10.1145/3038912.3052573>
- [51] USDA. 2023. *FoodData Central*. <https://fdc.nal.usda.gov/index.html>
- [52] Vaishali S. Vairale and Samiksha Shukla. 2021. Recommendation of food items for thyroid patients using content-based knn method. *Lecture Notes in Networks and Systems* 132 (2021), 71–77. [https://doi.org/10.1007/978-981-15-5309-7\\_8](https://doi.org/10.1007/978-981-15-5309-7_8)
- [53] Hongwei Wang, Fuzheng Zhang, Jialin Wang, Miao Zhao, Wenjie Li, Xing Xie, and Minyi Guo. 2018. RippleNet. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*. ACM, New York, NY, USA, 417–426. <https://doi.org/10.1145/3269206.3271739> arXiv:1803.03467
- [54] Wenjie Wang, Ling Yu Duan, Hao Jiang, Peiguang Jing, Xuemeng Song, and Liqiang Nie. 2021. Market2Dish: Health-aware Food Recommendation. *ACM Transactions on Multimedia Computing, Communications and Applications* 17, 1 (2021), 1–19. <https://doi.org/10.1145/3418211> arXiv:2012.06416
- [55] Xiting Wang, Yiru Chen, Jie Yang, Le Wu, Zhengtao Wu, and Xing Xie. 2018. A Reinforcement Learning Framework for Explainable Recommendation. In *2018 IEEE International Conference on Data Mining (ICDM)*, Vol. 2018-Novem. IEEE, 587–596. <https://doi.org/10.1109/ICDM.2018.00074>
- [56] Xiang Wang, Xiangnan He, Yixin Cao, Meng Liu, and Tat-Seng Chua. 2019. KGAT: Knowledge Graph Attention Network for Recommendation. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, New York, NY, USA, 950–958. <https://doi.org/10.1145/3292500.3330989> arXiv:1905.07854
- [57] Xiang Wang, Dingxian Wang, Canran Xu, Xiangnan He, Yixin Cao, and Tat-Seng Chua. 2019. Explainable Reasoning over Knowledge Graphs for Recommendation. *Proceedings of the AAAI Conference on Artificial Intelligence* 33 (jul 2019), 5329–5336. <https://doi.org/10.1609/aaai.v33i01.33015329> arXiv:1811.04540
- [58] Ingmar Weber and Palakorn Achananuparp. 2016. INSIGHTS FROM MACHINE-LEARNED DIET SUCCESS PREDICTION. *Pacific Symposium on Biocomputing, Pacific Symposium on Biocomputing* 21 (2016), 540–551.
- [59] WHO. 2015. *WHO calls on countries to reduce sugars intake among adults and children*. <https://www.who.int/news/item/04-03-2015-who-calls-on-countries-to-reduce-sugars-intake-among-adults-and-children>
- [60] WHO. 2023. *Obesity info page of the World Health Organization*. <https://www.who.int/health-topics/obesity>
- [61] David Wood, Richard Cyganiak, and Markus Lanthaler. 2014. *RDF 1.1 Concepts and Abstract Syntax*. W3C Recommendation. W3C. <https://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/>.
- [62] Yikun Xian, Zuohui Fu, Qiaoying Huang, S. Muthukrishnan, and Yongfeng Zhang. 2020. Neural-Symbolic Reasoning over Knowledge Graph for Multi-stage Explainable Recommendation. (jul 2020). arXiv:2007.13207 <http://arxiv.org/abs/2007.13207>
- [63] Yikun Xian, Zuohui Fu, S. Muthukrishnan, Gerard de Melo, and Yongfeng Zhang. 2019. Reinforcement Knowledge Graph Reasoning for Explainable Recommendation. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, New York, NY, USA, 285–294. <https://doi.org/10.1145/3331184.3331203> arXiv:1906.05237
- [64] Wenhan Xiong, Thien Hoang, and William Yang Wang. 2017. DeepPath: A Reinforcement Learning Method for Knowledge Graph Reasoning. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Stroudsburg, PA, USA, 564–573. <https://doi.org/10.18653/v1/D17-1060> arXiv:1707.06690
- [65] Shuyuan Xu, Yunqi Li, Shuchang Liu, Zuohui Fu, Xu Chen, and Yongfeng Zhang. 2020. Learning Post-Hoc Causal Explanations for Recommendation. *CEUR Workshop Proceedings 2911* (jun 2020), 13–25. arXiv:2006.16977 <http://arxiv.org/abs/2006.16977>
- [66] Raciél Yera Toledo, Ahmad A. Alzahrani, and Luis Martínez. 2019. A Food Recommender System Considering Nutritional Information and User Preferences. *IEEE Access* 7 (2019), 96695–96711. <https://doi.org/10.1109/ACCESS.2019.2929413>
- [67] Wenbin Yue, Zidong Wang, Jieyu Zhang, and Xiaohui Liu. 2021. An Overview of Recommendation Techniques and Their Applications in Healthcare. *IEEE/CAA Journal of Automatica Sinica* 8, 4 (apr 2021), 701–717. <https://doi.org/10.1109/JAS.2021.1003919>
- [68] Yongfeng Zhang and Xu Chen. 2020. Explainable Recommendation: A Survey and New Perspectives. *Foundations and Trends® in Information Retrieval* 14, 1 (2020), 1–101. <https://doi.org/10.1561/15000000066> arXiv:1804.11192
- [69] Michael Zimmermann. 1994. Diet, Nutrition, and the Prevention of Chronic Diseases. *The American Journal of Clinical Nutrition* 60, 4 (oct 1994), 644–645. <https://doi.org/10.1093/ajcn/60.4.644a>

## A ONLINE RESOURCES

All resources are available through our GitLab repository<sup>3</sup>.

<sup>3</sup><https://gitlab.com/felix134/connected-recipe-data-set>. The README.md file provides additional technical information regarding the data and steps to follow to reproduce the shown results, figures, and tables.