Norway grants

Modeling ingredient variations across contexts: Acquiring data and modeling knowledge for comprehensive gold and silver culinary recipe corpora

Agnieszka Ławrynowicz,
Associate Professor
Poznan University of Technology

The National Centre for Research and Development

What ingredient to substitute?

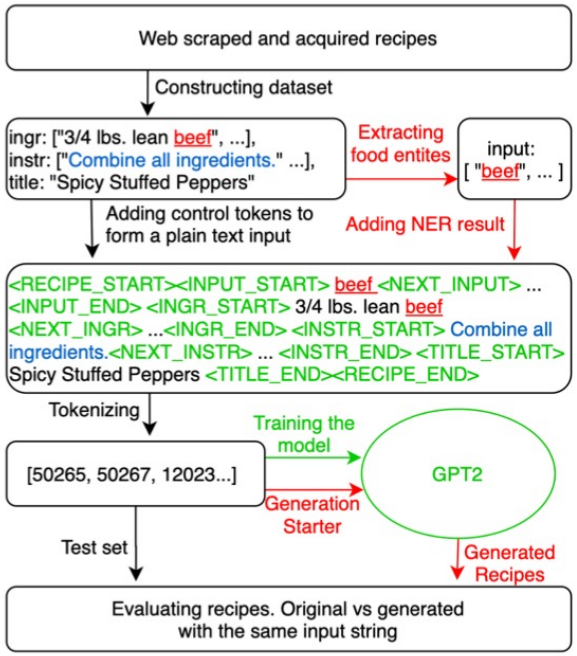
TAISTI is a project realised by a multi-disciplinary team developing AI-based technology for recommending ingredient substitutes in recipes.

Norway grants

The National Centre for Research and Development

Data: RecipeNLG

RecipeNLG dataset, partly based on the Recipe1M+ dataset, and providing over 1 million new, preprocessed and deduplicated recipes

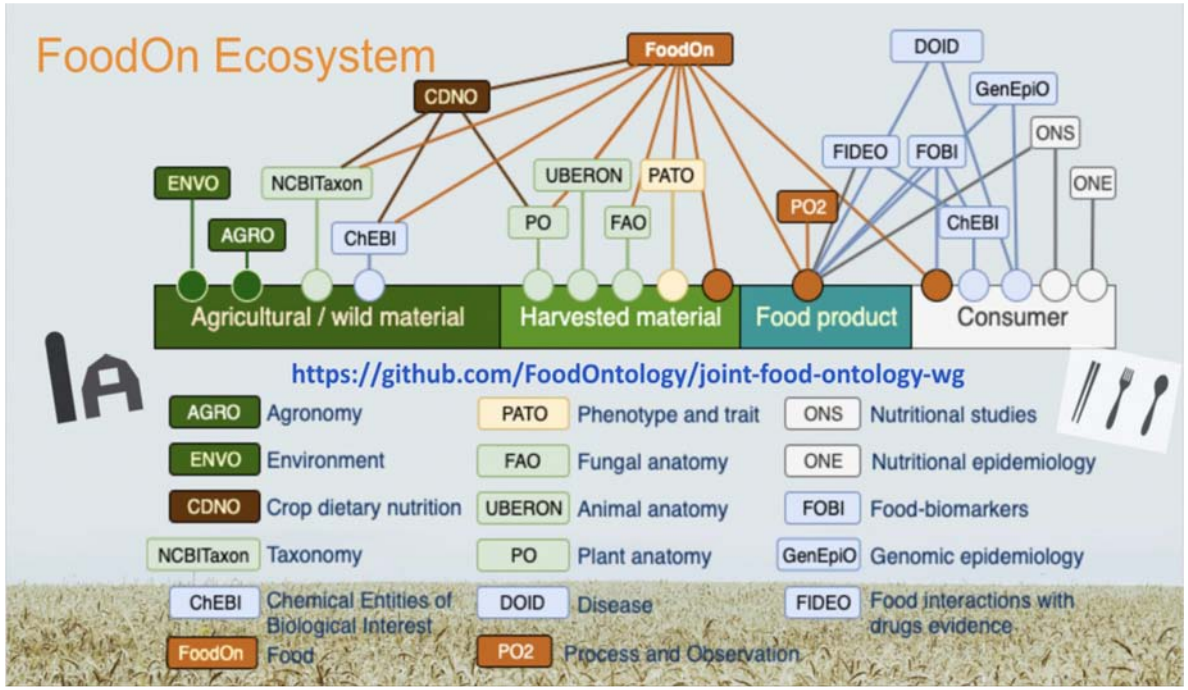


Michal Bien, Michal Gilski, Martyna Maciejewska, Wojciech Taisner, Dawid Wisniewski, Agnieszka Lawrynowicz, RecipeNLG: A Cooking Recipes Dataset for Semi-Structured Text Generation. INLG 2020: 22-28

Norway grants

The National Centre for Research and Development

Ontologies: FoodOn + PATO + ONS



Norway grants

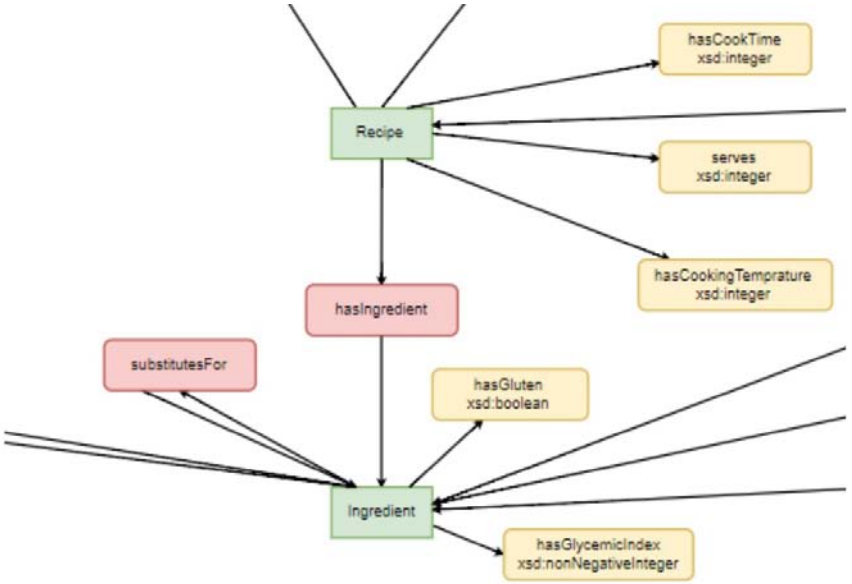
Dooley, D.M.; Griffiths, E.J.; Gosal, G.S.; Buttigieg, P.L.; Hoehndorf, R.; Lange, M.C.; Schriml, L.M.; Brinkman, F.S.L.; Hsiao, W.W.L. FoodOn: A harmonized food ontology to increase global food traceability, quality control and data integration. NPJ Sci. Food 2018, 2, 23.

The National Centre for Research and Development

Ontological models

What ingredient to substitute? Generic substitution

WhatToMake ontology
(HEALS project)



Norway grants

Sola S. Shirai, Oshani Seneviratne, Minor Gordon, Ching-Hua Chen, Deborah L. McGuinness: Identifying Ingredient Substitutions Using a Knowledge Graph of Food. Frontiers Artif. Intell. 3: 621766 (2020)

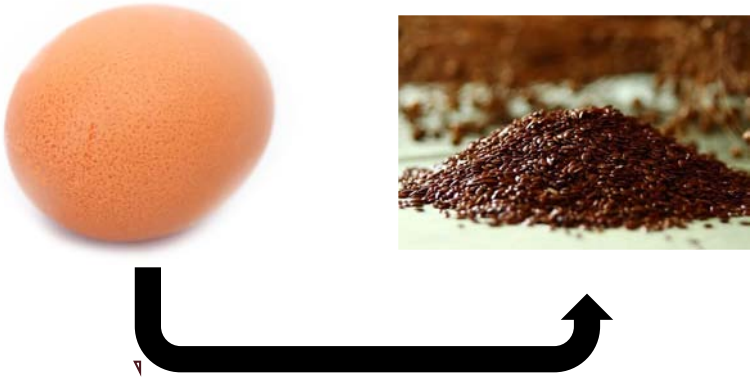
The National Centre for Research and Development

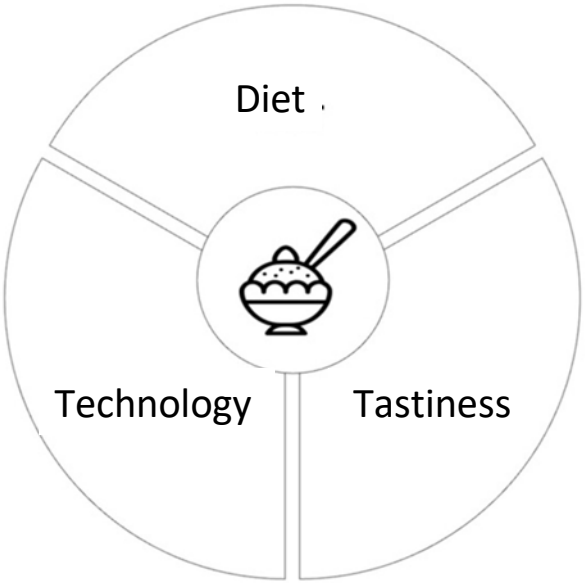
What ingredient to substitute, why and how?

Context:

Constraints: allergy, diet, health condition, lack of ingredient, …

Goals: exclude a specific product for health reasons, increase the intake of a particular nutrient, make a dish less dense, ...





Norway grants

The National Centre for Research and Development

Example substitution (recipe for pancakes)

Substitution
butter, melted—>regular butter Pre-processing step. Melt butter.
Ingredients
all-purpose flour—1½ cups salt—pinch or more to taste white sugar—1 tablespoon milk—1¼ cups egg—1 **butter—2½ tablespoons** oil for frying
Recipe process
Step 0. Melt butter. Step 1. Mix flour, salt, and sugar in a bowl. Add milk, egg, and melted butter. Step 2. Blend until smooth. Step 3. Heat a frying pan with light oil. Step 4. Pour the batter into the pan, using about 1/3 cup for each pancake. Step 5. Brown on both sides.

How to model substitutes for ingredients in food recipes?

- What related concepts should be taken into account to define the substitution's context (e.g., conditions, goals)?
- How to link the proposed model to existing food models and recommended design patterns?

Norway grants

The National Centre for Research and Development

Recipe ingredient substitute: binary or n-ary relation?

Conceptualization of substitutions in existing ontologies and knowledge graphs.

Ontology	Substitution	Context	Limitations
FoodOn	a (symmetric) relation: 'has food substance analog'	dietary and allergen analysis	no links with food preparation process, recipes
	subclasses of class: 'food product analog'		
FoodKG	heuristics based on explicit semantics and embeddings	dietary restrictions nutritional change	no ontological conceptualization
ONE	no term(s) for substitution		
ONS	no term(s) for substitution		

Norway grants

The National Centre for Research and Development

Recipe ingredient substitute: binary or n-ary relation?

Modeling ingredient substitutes - two major settings:

1. without context, i.e., as general substitutes
2. taking into account the context of a specific recipe and its food technological aspects and other factors, such as dietary constraints, goals, and technological conditions

In the latter case, relation expressing substitution may become *n*-ary

To model *n*-ary relation using OWL, we may reify it to a class **Substitution** and attach all „attributes" to it

Norway grants

The National Centre for Research and Development

Recipe ingredient substitution: process or specification?

- **Food recipe** specifies a procedure on how to prepare a dish
- It is not a **process** per se (spanning a time period) but rather a **(plan) specification** of a process to be performed to prepare a dish
- A major modeling decision: to model ingredient substitution in recipes on the level of **specification** rather than **processes**

Norway grants

The National Centre for Research and Development

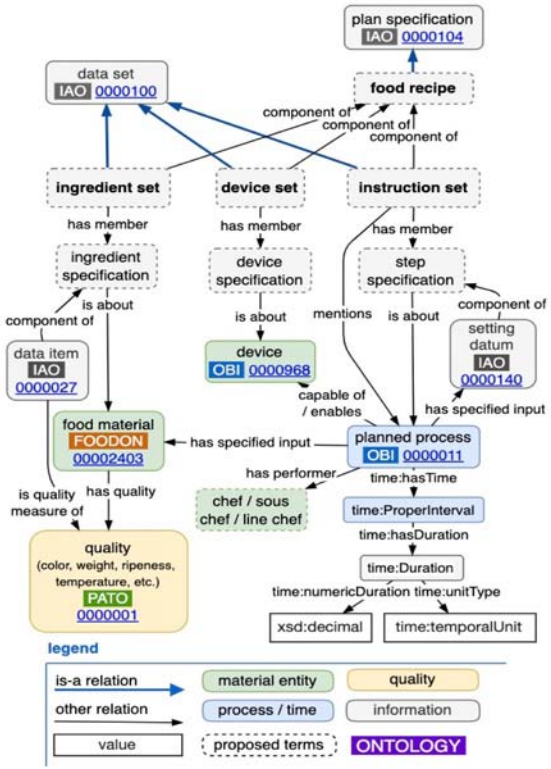
Ingredient specification

- A complex entity:
  - food items (materials)
  - quality measure of unit and quantity
- A simple one or a complex one:
  - disjunction of ingredients (**goat or sheep meat or both**)
  - conjunction (**dill and fennel**, **green and dry onions**) or combinations of spices)
  - optional -> ’empty’ substitution?

Norway grants

The National Centre for Research and Development

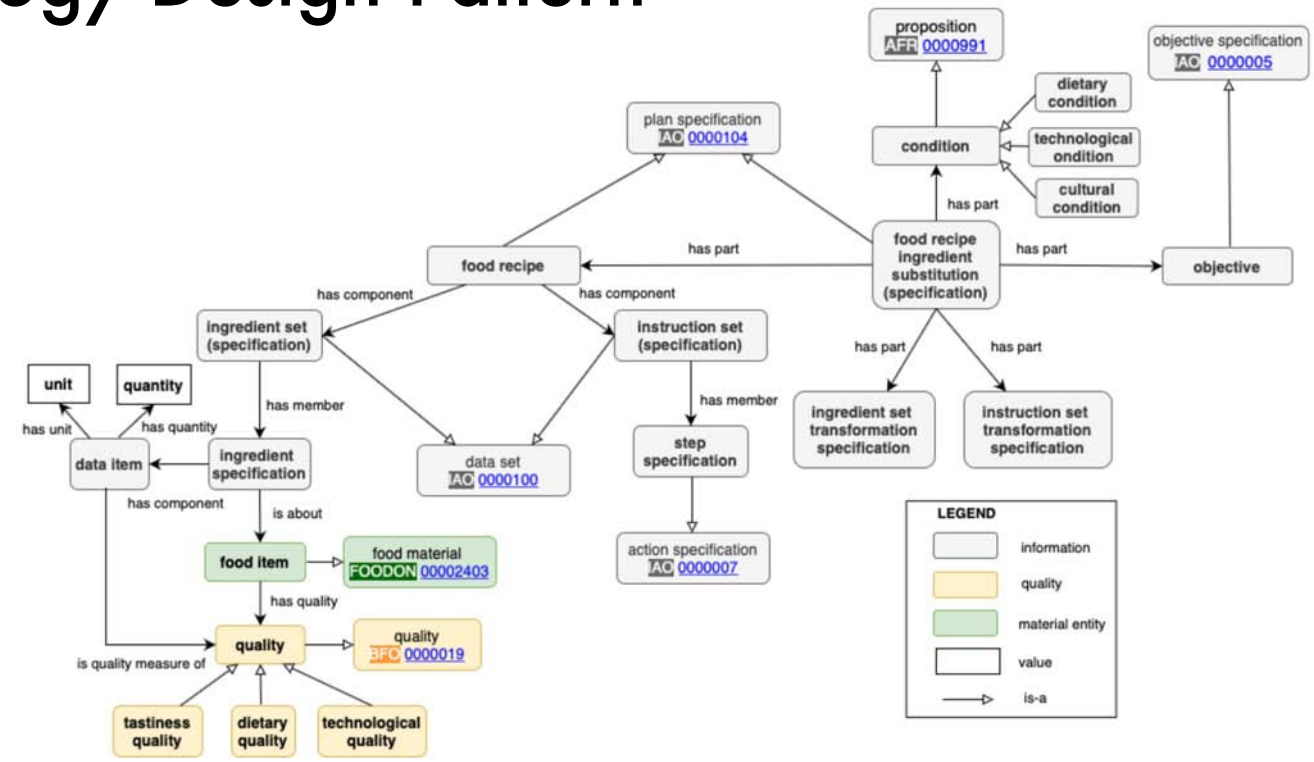
FoodOn recipe model



Norway grants

Lange, M.; Dooley, D.; Weber, M.; Ibanescu, L.; Chan, L.; Soldatova, L.; Yang, C.; Warren, R.; Shimizu, C.; McGinty, H.K.; et al. Food Process Ontology Requirements. Available online: http://www.semantic-web-journal.net/content/food-process-ontology-requirements (accessed on 15 December 2021).

Food Recipe Ingredient Substitution Ontology Design Pattern



Norway grants

Agnieszka Ławrynowicz, Anna Wróblewska, Weronika T. Adrian, Bartosz Kulczynski, Anna Gramza-Michalowska. Food Recipe Ingredient Substitution Ontology Design Pattern. Sensors 22(3): 1095 (2022)

Is this the only right solution?

What about finding another (binary?) role/relationship that connects two things that can substitute one another so they have the same role to products instead of „has food product analog”?

Norway grants

The National Centre for Research and Development

Modeling substitutes - FoodOn curation workgroup*

- **food product //OR// food by quality**
  - **milk food product**
  - **non-diary milk food product**
    - **plant milk product**
      - **nut milk product**
      - **…**

Norway grants

*joint modeling effort of FoodOn curation workgroup (with Damion Dooley, Rhiannon Cameron, Magalie Weber, Kai Blumberg, other FoodOn curators)

The National Centre for Research and Development

- For example natural or artificial „vanilla”: differences in provenance of the material, but also differences in organoleptics
- Let’s think about „vanilla taste” role
- Both could then have the same aromatic role

Norway grants

*joint modeling effort of FoodOn curation workgroup (with Damion Dooley, Rhiannon Cameron, Magalie Weber, Kai Blumberg, other FoodOn curators)

The National Centre for Research and Development

What about using „quality”?

E.g. „milk quality”?

But then we may have:

„milk quality” with respect to taste

„milk quality” with respect to nutritional value?

Norway grants

*joint modeling effort of FoodOn curation workgroup (with Damion Dooley, Rhiannon Cameron, Magalie Weber, Kai Blumberg, other FoodOn curators)

The National Centre for Research and Development

What about using „quality”?

E.g. „milk quality”?

But then we may have:

„milk quality” with respect to taste

„milk quality” with respect to nutritional value?

„milk organoleptic quality”, „milk fat quality”, …

Norway grants

*joint modeling effort of FoodOn curation workgroup (with Damion Dooley, Rhiannon Cameron, Magalie Weber, Kai Blumberg, other FoodOn curators)

The National Centre for Research and Development

Named entity recognition models

TASTEset - Recipe Dataset and Food Entities Recognition Benchmark

Tagset:

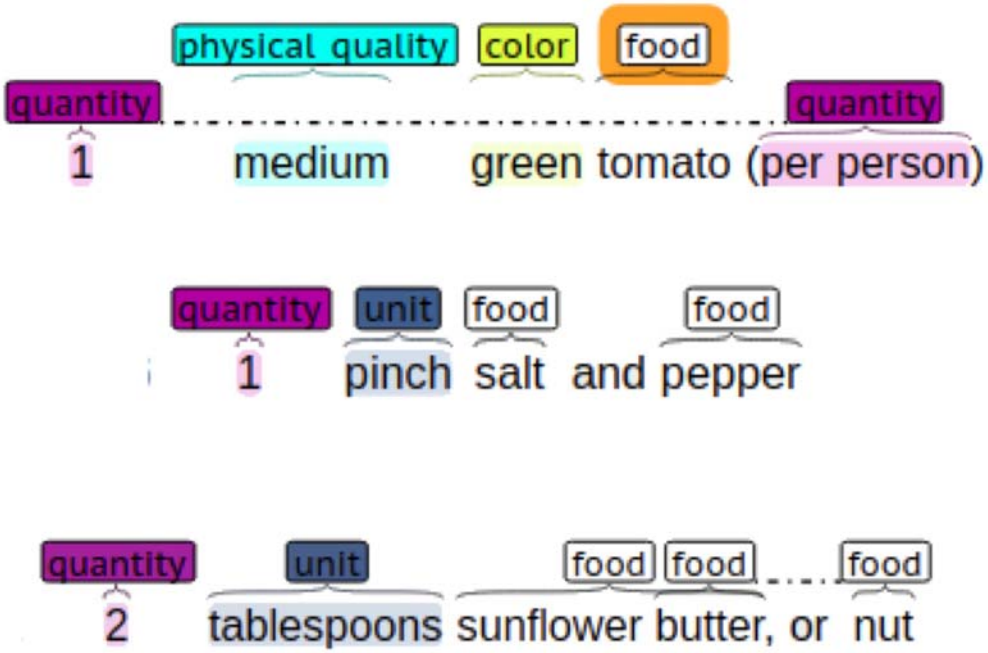
- FOOD as the name of an ingredient (e.g. bread, mayonnaise, salt, tomato),
- QUANTITY as a quantity (usually expressed by digits or a float),
- UNIT as a unit of measurement (e.g. bunch, cup, grams, jar, millimeters, slices, stalks, tablespoon, teaspoon)
- PROCESS as the attribute of the ingredient, usually referring to an action to be taken to prepare the ingredient (e.g. chopped for parsley, crushed for garlic, grated for ginger, minced for garlic),
- PHYSICAL QUALITY as the characteristic of the ingredient (e.g. boneless for chicken breast),
- COLOR as the color of the ingredient,
- TASTE as the flavour (e.g. bittersweet, butter-flavoured, sweet, semi-sweet),
- PURPOSE as the purpose of using the ingredient in the recipe (e.g. for dusting about flour, for garnish about sunflower seeds, for frying about canola oil, and as topping about sour cream),
- PART as a part of the ingredient required by the recipe (e.g. yolks and whites as parts of the eggs).

Norway grants

Anna Wróblewska, Agnieszka Kaliska, Maciej Pawlowski, Dawid Wisniewski, Witold Sosnowski, Agnieszka Lawrynowicz: TASTEset - Recipe Dataset and Food Entities Recognition Benchmark. CoRR abs/2204.07775 (2022) https://arxiv.org/abs/2204.07775

The National Centre for Research and Development

TASTEset: entity examples



Norway grants

The National Centre for Research and Development

TASTEset

- 700 recipes with more than 13,000 entities to extract
- manual annotation covered 3,788 ingredients of varying complexity
- state-of-the-art baselines of named entity recognition models

List of ingredients
- 1 box powdered sugar
- 8 oz. soft butter
- 1 (8 oz.) peanut butter
- paraffin
- 12 oz. chocolate chips

1 QUANTITY box UNIT powdered PHYSICAL_QUALITY sugar FOOD
8 QUANTITY oz UNIT . soft PHYSICAL_QUALITY butter FOOD
1 QUANTITY (8 QUANTITY oz UNIT .) peanut butter FOOD
paraffin FOOD
12 QUANTITY oz UNIT . chocolate chips FOOD

Norway grants

The National Centre for Research and Development

Named Entity Recognition - results

entity	Model				
	$\text{BERT}_{\text{base-cased}}$	$\text{BERT}_{\text{large-cased}}$	FoodNER	BERT with CRF	LUKE
FOOD	0.889 ± 0.013	0.898 ± 0.012	0.888 ± 0.010	0.895 ± 0.013	$\mathbf{0.915 \pm 0.005}$
QUANTITY	0.983 ± 0.004	0.982 ± 0.007	$\mathbf{0.985 \pm 0.004}$	0.980 ± 0.006	0.975 ± 0.010
UNIT	$\mathbf{0.979 \pm 0.005}$	0.976 ± 0.009	$\mathbf{0.979 \pm 0.005}$	0.975 ± 0.006	0.976 ± 0.006
PROCESS	$\mathbf{0.930 \pm 0.006}$	0.927 ± 0.015	0.921 ± 0.007	0.913 ± 0.019	0.916 ± 0.021
PHYSICAL QUALITY	0.797 ± 0.011	0.795 ± 0.025	$\mathbf{0.804 \pm 0.015}$	0.772 ± 0.003	0.767 ± 0.035
COLOR	0.873 ± 0.039	0.910 ± 0.039	$\mathbf{0.913 \pm 0.037}$	0.894 ± 0.046	0.872 ± 0.051
TASTE	0.772 ± 0.115	0.781 ± 0.089	0.769 ± 0.116	$\mathbf{0.789 \pm 0.093}$	0.655 ± 0.120
PURPOSE	0.789 ± 0.145	$\mathbf{0.853 \pm 0.141}$	0.801 ± 0.140	0.821 ± 0.127	0.684 ± 0.176
PART	0.705 ± 0.115	$\mathbf{0.801 \pm 0.034}$	0.744 ± 0.127	0.730 ± 0.091	0.786 ± 0.10
all	0.932 ± 0.008	$\mathbf{0.935 \pm 0.011}$	0.932 ± 0.006	0.929 ± 0.008	0.927 ± 0.008

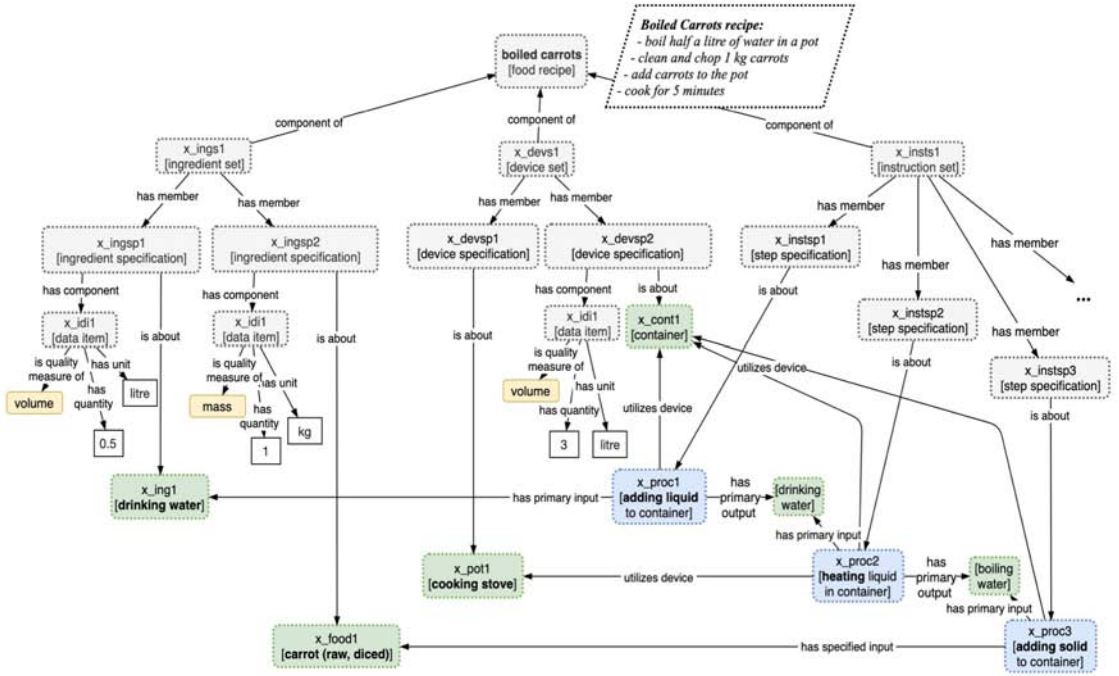
Table 2: The detailed results (average F_1-scores over 5 runs in 5-fold cross-validation) of our baselines for *TASTEset* dataset.

Norway
grants

The National Centre
for Research and Development

Knowledge graph construction

FoodOn recipe model - example

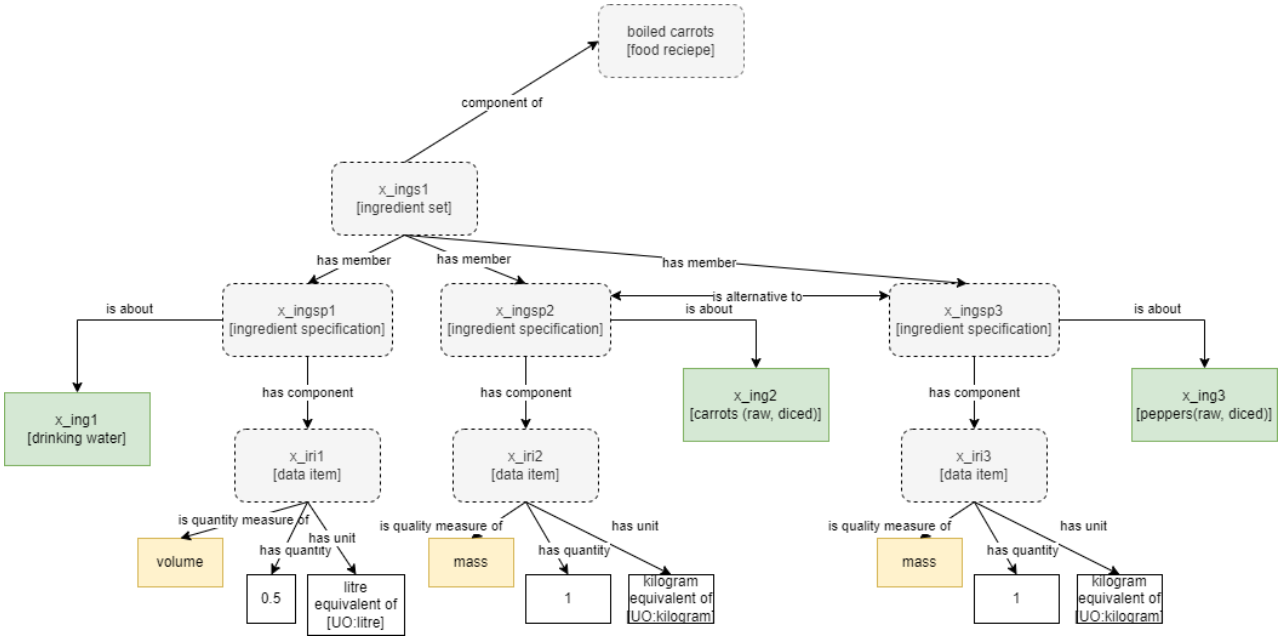


Norway grants

Lange, M.; Dooley, D.; Weber, M.; Ibanescu, L.; Chan, L.; Soldatova, L.; Yang, C.; Warren, R.; Shimizu, C.; McGinty, H.K.; et al. Food Process Ontology Requirements. Available online: http://www.semantic-web-journal.net/content/food-process-ontology-requirements (accessed on 15 December 2021).

The National Centre for Research and Development

Modeling alternative ingredients



Norway grants

The National Centre for Research and Development

What instances (not) to create

OWL has been designed to distinguish between a class and an instance

Consider: **carrots (cooked, diced)** and that we have classes **carrot food product**, **cooked** (quality)**, diced** (quality)

Should we thus create in a knowledge graph new instances for each „attribute” of an ingredient (e.g. for **sweet** and **diced**)?

It would be inefficient to copy these „classes” with just one new instance each time

What instances (not) to create: punning

OWL designed to distinguish between a class and an instance

Solution: **punning**

Norway grants

The National Centre for Research and Development

Entity linking - LexMapr

green onions (chopped)

['onion (chopped):FOODON_03316397', 'scallion (whole, raw):FOODON_03311340'] Component Match

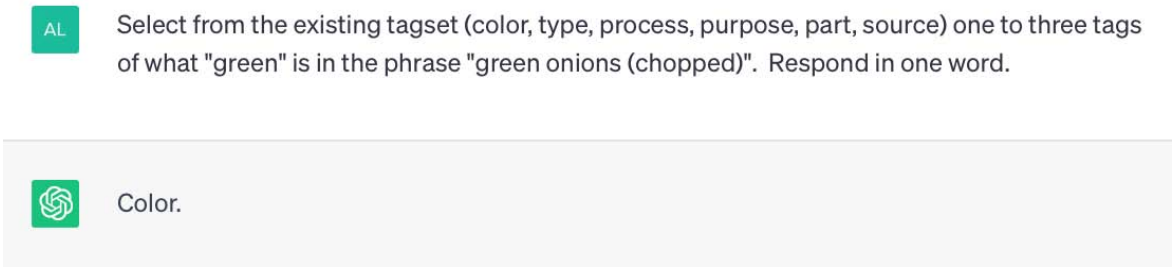
Norway grants

The National Centre for Research and Development

Could we ask LLM for adding missing information?

green onions (chopped)

Use existing tags in a prompt

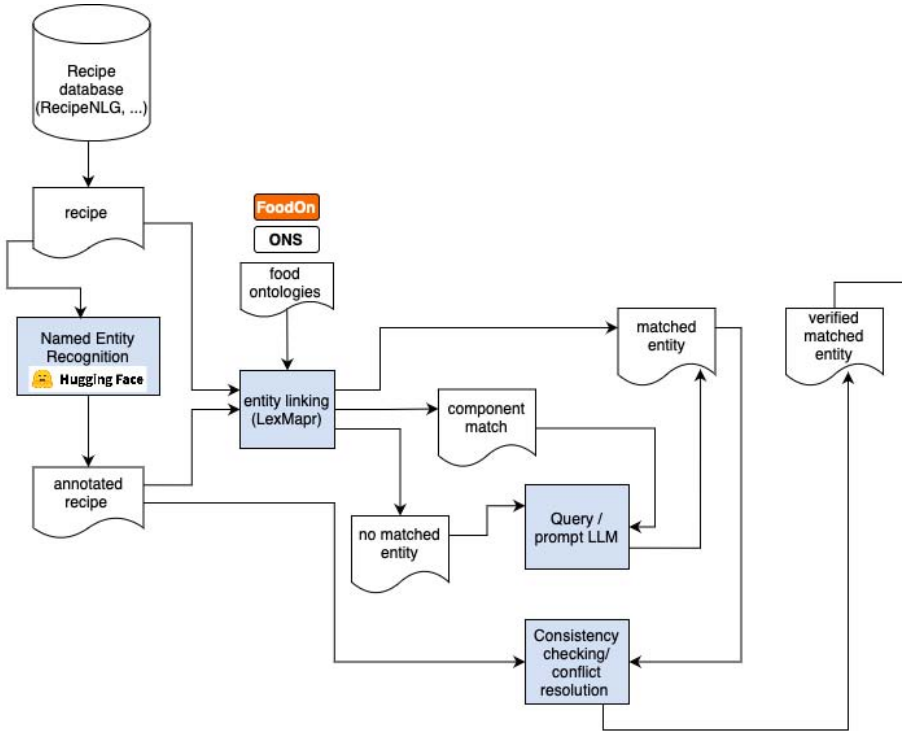


Norway grants

The National Centre for Research and Development

Entity linking with LLM and logical reasoning

Active labeling (augmentation) using human annotators as „strong experts” (gold coprora) and LLM as „weak expert” (silver copora)

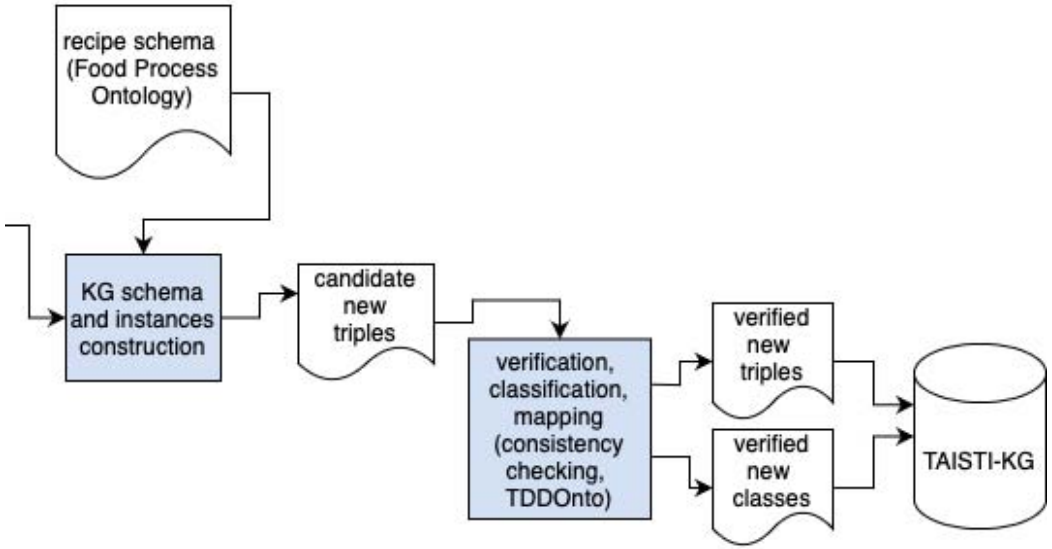


Norway grants

The National Centre for Research and Development

Ontology / KG authoring

Logic based Test-Driven Development of ontologies/knowledge graphs



Norway grants

The National Centre for Research and Development

TAISTI team

Agnieszka Ławrynowicz

Anna Wróblewska

Weronika T. Adrian

Dawid Wiśniewski

Agnieszka Kaliska

Anna Gramza-Michałowska

and
Jędrzej Potoniec, Mateusz Lango, Witold Sosnowski, Maciej Pawłowski, Bartosz Kulczyński, Sergiy Tkachuk, Adam Lewandowski, Andrzej Gretkowski, Maciej Kutyła, Jakub Dutkiewicz

Norway grants

The National Centre for Research and Development