

# Semantic Web – Food

# **The Knowledge Graph**

Aspir Ahmet

Dick Tobias

Egger Daniel

Kaufmann Stefan

# Overview

- Summary of first Report
  - Usecase
  - Domain Analysis
  - Dataset Descriptions
  - Vocabulary
  - Knowledge Graph Statistics
- Data Enrichment
  - LOD Linking
  - Reasoner Results
  - SHACL
  - Information Extraction (NLP)
- Conclusion and Future Work

# Usecase

- Knowledge-graph recipe application
- Recipe search on a Website:
  - ingredients
  - cook time
  - health-labels
  - ...
- Multifunctional search with respect to different user-intentions

# Domain Analysis

- Food  $\Leftrightarrow$  Recipe
- Core elements:
  - Ingredients
  - Instructions
  - Calories
  - Prepare-Time
- Miscellaneous:
  - Example Image
  - Author
  - ....

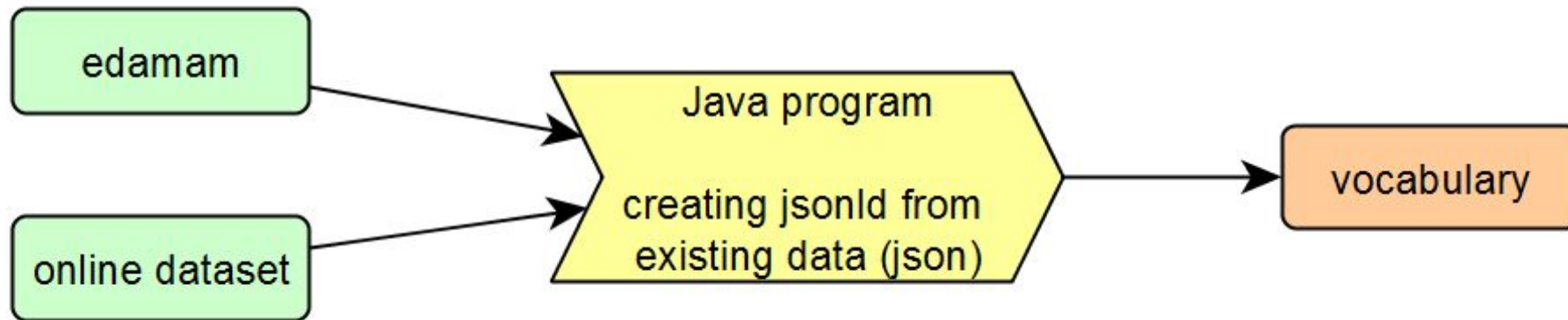
# Domain Model

- **Edamam API**

- 1,7+ million recipes

- **Online Dataset**

- 518 recipes



- **ReciPal**

- **Recipepuppy**

# Dataset Descriptions

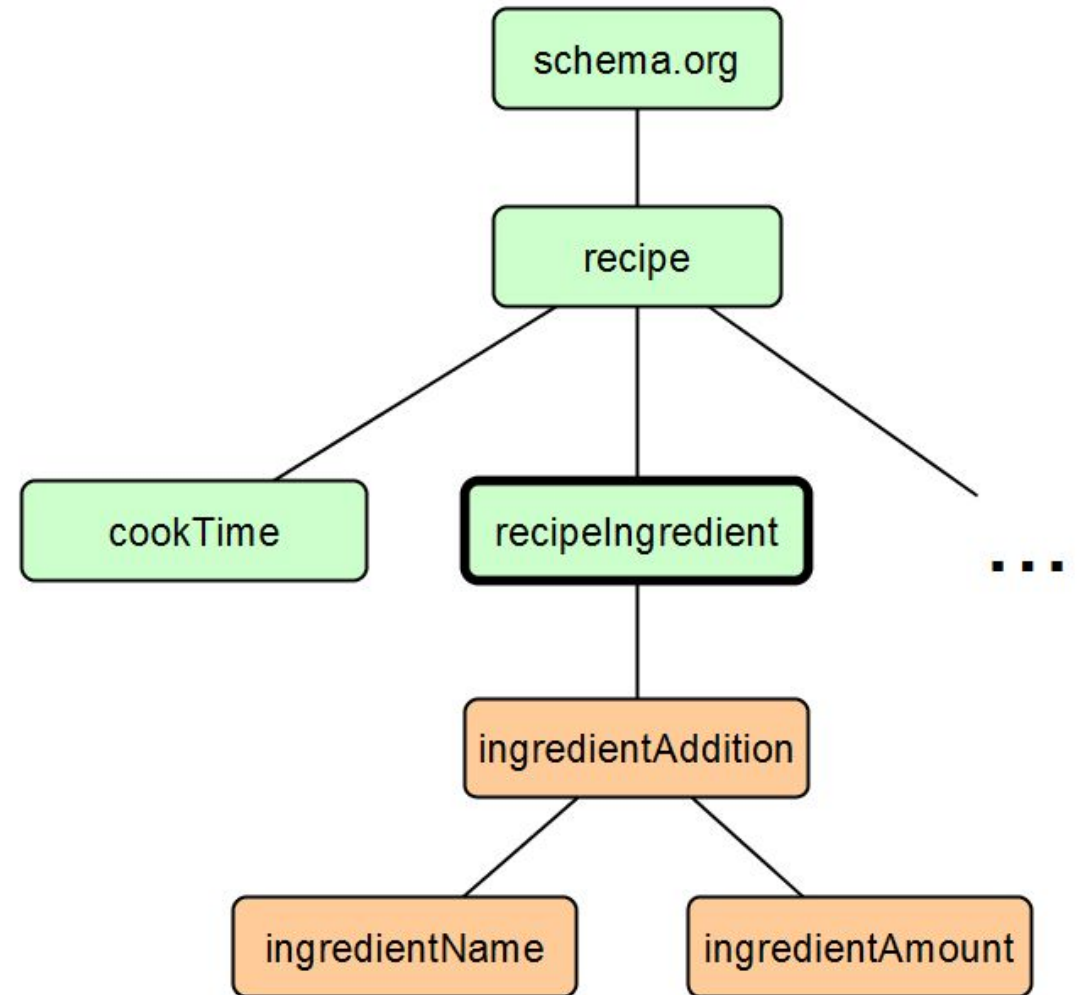
- Biggest and main supplier of data: EDAMAM
- API returns json file
- Is the richest in regards to information



Field	Type	Description
<code>uri</code>	<b>string</b>	Ontology identifier
<code>label</code>	<b>string</b>	Recipe title
<code>image</code>	<b>string</b>	Image URL
<code>source</code>	<b>string</b>	Source site identifier
<code>url</code>	<b>string</b>	Original recipe URL
<code>yield</code>	<b>integer</b>	Number of servings
<code>calories</code>	<b>float</b>	Total energy, kcal
<code>totalWeight</code>	<b>float</b>	Total weight, g
<code>ingredients</code>	<b>Ingredient[]</b>	Ingredients list
<code>totalNutrients</code>	<b>NutrientInfo[]</b>	Total nutrients for the entire recipe
<code>totalDaily</code>	<b>NutrientInfo[]</b>	% daily value for the entire recipe
<code>dietLabels</code>	<b>enum[]</b>	Diet labels: "balanced", "high-protein", "high-fiber", "low-fat", "low-carb", "low-sodium" (labels are per serving)
<code>healthLabels</code>	<b>enum[]</b>	Health labels: "vegan", "vegetarian", "paleo", "dairy-free", "gluten-free", "wheat-free", "fat-free", "low-sugar", "egg-free", "peanut-free", "tree-nut-free", "soy-free", "fish-free", "shellfish-free" (labels are per serving)

# Vocabulary

- schema.org
  - Base Vocabulary
- extension of recipeIngredient
  - Adding granularity to ingredients



# Knowledge Graph Statistics

- Total Number of Triples: 479899
- Number of distinct Classes: 8
- Number of distinct Properties: 19



# Knowledge Graph Statistics

## Classes

## Instances

< <a href="http://schema.org/IngredientAddition">http://schema.org/IngredientAddition</a> >	32465
< <a href="http://schema.org/CreativeWork">http://schema.org/CreativeWork</a> >	4433
< <a href="http://schema.org/NutritionInformation">http://schema.org/NutritionInformation</a> >	4433
< <a href="http://schema.org/Person">http://schema.org/Person</a> >	4433
< <a href="http://schema.org/Recipe">http://schema.org/Recipe</a> >	4433
< <a href="http://schema.org/Ingredient">http://schema.org/Ingredient</a> >	32465
< <a href="http://schema.org/QuantitativeValue">http://schema.org/QuantitativeValue</a> >	36898
< <a href="http://schema.org/ImageObject">http://schema.org/ImageObject</a> >	4433

# Knowledge Graph Statistics

Properties	Subjects	Objects
<a href="http://schema.org/calories">&lt;http://schema.org/calories&gt;</a>	4433	2891
<a href="http://schema.org/unitText">&lt;http://schema.org/unitText&gt;</a>	32465	765
<a href="http://schema.org/ingredientName">&lt;http://schema.org/ingredientName&gt;</a>	32465	32465
<a href="http://schema.org/totalTime">&lt;http://schema.org/totalTime&gt;</a>	4433	361
<a href="http://schema.org/prepTime">&lt;http://schema.org/prepTime&gt;</a>	4433	1
<a href="http://schema.org/recipeYield">&lt;http://schema.org/recipeYield&gt;</a>	4433	4433
<a href="http://schema.org/image">&lt;http://schema.org/image&gt;</a>	4433	4433
<a href="http://schema.org/contentUrl">&lt;http://schema.org/contentUrl&gt;</a>	4433	4334
<a href="http://schema.org/name">&lt;http://schema.org/name&gt;</a>	41331	18974
<a href="http://schema.org/cookTime">&lt;http://schema.org/cookTime&gt;</a>	4433	1
<a href="http://schema.org/nutrition">&lt;http://schema.org/nutrition&gt;</a>	4433	4433
<a href="http://schema.org/creator">&lt;http://schema.org/creator&gt;</a>	4433	4433
<a href="http://schema.org/recipeIngredient">&lt;http://schema.org/recipeIngredient&gt;</a>	4433	38233
<a href="http://schema.org/url">&lt;http://schema.org/url&gt;</a>	4433	4346
<a href="http://schema.org/recipeInstructions">&lt;http://schema.org/recipeInstructions&gt;</a>	4433	4433
<a href="http://schema.org/caption">&lt;http://schema.org/caption&gt;</a>	4433	3960
<a href="http://schema.org/value">&lt;http://schema.org/value&gt;</a>	36898	354
<a href="http://schema.org/sameAs">&lt;http://schema.org/sameAs&gt;</a>	4433	4346
<a href="http://www.w3.org/1999/02/22-rdf-syntax-ns#type">&lt;http://www.w3.org/1999/02/22-rdf-syntax-ns#type&gt;</a>	123993	8

# Knowledge Graph Statistics

- Properties in Top5 Classes

## Property

[<http://schema.org/unitText>](http://schema.org/unitText)  
[<http://schema.org/value>](http://schema.org/value)  
[<http://schema.org/ingredientFullName>](http://schema.org/ingredientFullName)  
[<http://schema.org/name>](http://schema.org/name)  
[<http://schema.org/ingredientName>](http://schema.org/ingredientName)  
[<http://schema.org/ingredientAmount>](http://schema.org/ingredientAmount)  
[<http://schema.org/url>](http://schema.org/url)  
[<http://schema.org/contentUrl>](http://schema.org/contentUrl)  
[<http://schema.org/caption>](http://schema.org/caption)

## Class

[<http://schema.org/QuantitativeValue>](http://schema.org/QuantitativeValue)  
[<http://schema.org/QuantitativeValue>](http://schema.org/QuantitativeValue)  
[<http://schema.org/Ingredient>](http://schema.org/Ingredient)  
[<http://schema.org/Ingredient>](http://schema.org/Ingredient)  
[<http://schema.org/IngredientAddition>](http://schema.org/IngredientAddition)  
[<http://schema.org/IngredientAddition>](http://schema.org/IngredientAddition)  
[<http://schema.org/CreativeWork>](http://schema.org/CreativeWork)  
[<http://schema.org/ImageObject>](http://schema.org/ImageObject)  
[<http://schema.org/ImageObject>](http://schema.org/ImageObject)

# Knowledge Graph Statistics

- Data Alignment

Wikidata Item	Wikilabel	Localclass
< <a href="http://www.wikidata.org/entity/Q7759779">http://www.wikidata.org/entity/Q7759779</a> >	"The Recipe"@en	< <a href="http://schema.org/Recipe">http://schema.org/Recipe</a> >
< <a href="http://www.wikidata.org/entity/Q11241662">http://www.wikidata.org/entity/Q11241662</a> >	"Recipe"@en	< <a href="http://schema.org/Recipe">http://schema.org/Recipe</a> >
< <a href="http://www.wikidata.org/entity/Q21188738">http://www.wikidata.org/entity/Q21188738</a> >	"Recipe"@en	< <a href="http://schema.org/Recipe">http://schema.org/Recipe</a> >
< <a href="http://www.wikidata.org/entity/Q219239">http://www.wikidata.org/entity/Q219239</a> >	"recipe"@en	< <a href="http://schema.org/Recipe">http://schema.org/Recipe</a> >
< <a href="http://www.wikidata.org/entity/Q605076">http://www.wikidata.org/entity/Q605076</a> >	"cookbook"@en	< <a href="http://schema.org/Recipe">http://schema.org/Recipe</a> >
< <a href="http://www.wikidata.org/entity/Q605076">http://www.wikidata.org/entity/Q605076</a> >	"Cookbook"@en-ca	< <a href="http://schema.org/Recipe">http://schema.org/Recipe</a> >
< <a href="http://www.wikidata.org/entity/Q605076">http://www.wikidata.org/entity/Q605076</a> >	"cookery book"@en-gb	< <a href="http://schema.org/Recipe">http://schema.org/Recipe</a> >

# LOD Linking

- creation of an entry
  - specify the wikidata-IRI
- name the entity the same as the Q identifier from wikidata (Recipe = Q219239)

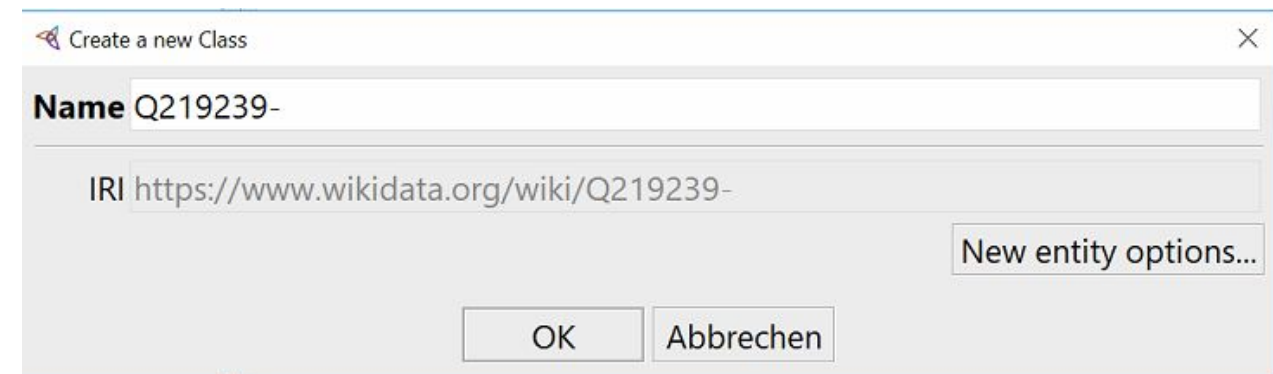


Entity Creation Preferences

**Entity IRI**

Start with: ☐ Active ontology IRI  
☒ Specified IRI:

Followed by: ☒ #  
☐ /  
☐ :



Create a new Class

**Name**

**IRI**

# LOD Linking

- imported the schema-vocabulary to Protege
- extended it with our own vocab-extensions (schema:Food) —
- stated OWL axioms for the Recipe class —

The screenshot displays the Protege ontology editor interface. On the left, a hierarchical tree of classes is shown, with 'Recipe' and 'schema:Food' highlighted by a red box. The main panel on the right shows the 'Description: Recipe' tab. It lists several axioms for the 'Recipe' class, which are highlighted by a green box:

- Equivalent To: Q219239
- SubClass Of: Action, HowTo
- not (schema:Food)
- schema:hasImage max 1 owl:Thing
- schema:hasIngredient min 3 owl:Thing
- schema:hasRecipeInstructions some CreativeWork

Below these axioms, the 'General class axioms' section shows 'Recipe' as a subclass of an anonymous ancestor. The 'Instances' section shows 'SchemaRecipe' as an instance of 'Recipe'.

# LOD Linking

- wikidata entity of recipe is set to equal with our recipe from schema
- 
- subclasses are automatically inferred by the equivalence to the schema recipe entry
- 

The screenshot displays the Wikidata class page for **Q219239**, which is **Recipe**. The page is organized into several sections, each with a plus icon to expand it. A green box highlights the **Description** and **Equivalent To** sections, while a red box highlights the **SubClass Of (Anonymous Ancestor)** section.

**Description:** Q219239

**Equivalent To:**

- Recipe

**SubClass Of:**

- Action
- HowTo

**General class axioms:**

**SubClass Of (Anonymous Ancestor):**

- schema:hasImage **max** 1 owl:Thing
- schema:hasIngredient **min** 3 owl:Thing
- schema:hasRecipeInstructions **some** CreativeWork
- not** (schema:Food)
- Q219239

**Instances:**

**Target for Key:**

**Disjoint With:**

**Disjoint Union Of:**

# LOD Linking - Reasoner Results

- reasoner HermiT  
1.3.8.413
- explanation of the  
reasoner why  
something is inferred

The screenshot shows a list of constraints for Q219239:

- schema:hasIngredient **min** 3 owl:Thing
- schema:hasRecipeInstructions **some** CreativeWork
- **not** (schema:Food)
- Q219239

Below the constraints is a section titled "Instances" with a plus icon. It contains two items:

- ◆ Recipe
- ◆ SchemaRecipe

A mouse cursor is hovering over "SchemaRecipe", and a blue box labeled "Inferred" is positioned next to it. Below the instances is a section titled "Target for Key" with a plus icon.

Explanation for: Recipe Type Q219239

SchemaRecipe **SameAs** Recipe

SchemaRecipe **Type** Recipe

Recipe **EquivalentTo** Q219239



# SHACL

- used for data type and class validation
- restrict amount of recipe ingredients from 2 - 15

```
@prefix schema: <http://schema.org/> .  
@prefix sh: <http://www.w3.org/ns/shacl#> .  
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
```

```
schema:RecipeShape  
  a sh:NodeShape ;  
  sh:targetClass schema:Recipe ;  
  sh:property [  
    sh:path schema:name ;  
    sh:datatype xsd:string ;  
  ] ;  
  sh:property [  
    sh:path schema:creator ;  
    sh:class schema:Person ;  
  ] ;  
  sh:property [  
    sh:path schema:recipeYield ;  
    sh:class schema:QuantitativeValue ;  
  ] ;  
  sh:property [  
    sh:path schema:recipeInstructions ;  
    sh:class schema:CreativeWork ;  
  ] ;  
  sh:property [  
    sh:path schema:recipeIngredient ;  
    sh:class schema:IngredientAddition ;  
    sh:maxCount 15 ;  
    sh:minCount 2 ;  
  ] .
```

# SHACL

## Data Graph

```
"recipeIngredient": [  
  {  
    "@type" : "IngredientAddition",  
    "ingredientName" : {  
      "@type" : "Ingredient",  
      "name" : "Mustard",  
      "ingridientFullName" : "Dijon"  
    },  
    "ingredientAmount" : {  
      "@type" : "QuantitativeValue",  
      "unitText" : "Tablespoons",  
      "value" : "3"  
    }  
  }  
]
```

## Shapes Graph

```
sh:property [  
  sh:path schema:recipeIngredient ;  
  sh:class schema:IngredientAddition ;  
  sh:maxCount 15 ;  
  sh:minCount 2 ;  
] .
```

## Validation Report

```
[  
  a sh:ValidationResult ;  
  sh:resultSeverity sh:Violation ;  
  sh:sourceConstraintComponent sh:MinCountConstraintComponent ;  
  sh:sourceShape _:n1463 ;  
  sh:focusNode _:n1594 ;  
  sh:resultPath schema:recipeIngredient ;  
  sh:resultMessage "Less than 2 values" ;  
]
```

# Information Extraction (NLP)

Ingredient parsing is hard

1 glass red wine  
2.0 tbsp vegetable oil or beef fat  
4 eggs  
10 curry leaves  
1 yellow pepper, deseeded and thinly sliced  
1-1.5kg/2lb 4-3lb 5oz sirloin of beef joint  
1 Whole Beef Tenderloin (3 ½ to 5 Pounds)

Naive Regex Parser Examples

# Information Extraction (NLP)

Use ML (NER) instead


1	QUANTITY
glass	UNIT
red	NAME
wine	NAME

1	QUANTITY
yellow	NAME
pepper	NAME
,	0
deseeded	0
and	0
thinly	0
sliced	0

...



Training



Token  
Classifier  
Model

# Information Extraction (NLP)

## Results

### Naive Regex Parser

~**86.2%** of lines parseable  
unknown error rating  
obviously pretty bad  
convoluted and hard to maintain

### NER

after training with ~**5000** examples  
and testing with ~**1000** examples

~**93.7%** of tokens correctly labelled (recall)  
~**94.9%** of labels correct (precision)  
=>  $F_1$  score: ~**94.3%**

training progress indicates even more  
potential with more data

# Conclusion and Future Work

- add actions
- add second vocabulary
- extend dataset
- develop application

The screenshot shows a web browser window with a single tab titled 'Neuer Tab'. The address bar is empty, and the search bar contains the text 'Suchen'. The main content area features a list of filter options, each with a checkbox and a label:

- ☒ vegetarian
- ☒ low-carb
- ☐ ...
- ☐ ...
- ☒ total time < 30 minutes
- ☐ show recipes which only include given ingredients

Below the filters is a text input field with the placeholder text '<enter ingredients>' and a right-pointing arrow button.

At the bottom, there are three recipe cards, each with a title, a 'buy ingr.' button, and a list of ingredients:

Recipe 1	Recipe 2	Recipe 3
<b>Ingredients:</b> <ul style="list-style-type: none"><li>- 3 carrots</li><li>- 3 onions</li><li>- salt</li><li>- pepper</li></ul>	<b>Ingredients:</b> <ul style="list-style-type: none"><li>- 3 carrots</li><li>- 3 onions</li><li>- salt</li><li>- pepper</li></ul>	<b>Ingredients:</b> <ul style="list-style-type: none"><li>- 3 carrots</li><li>- 3 onions</li><li>- salt</li><li>- pepper</li></ul>