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

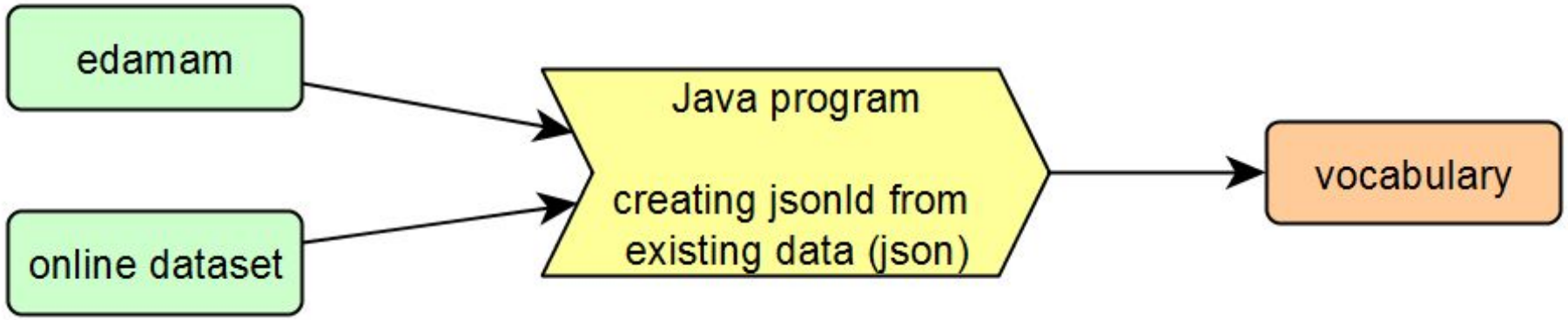
- Knowledge-graph recipe application
- Recipe search on a Website:
  - ingredients
  - cook time
  - health-labels
  - ...
- Multifunctional search with respect to different user-intentions

Domain Analysis

- Food ⇔ 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

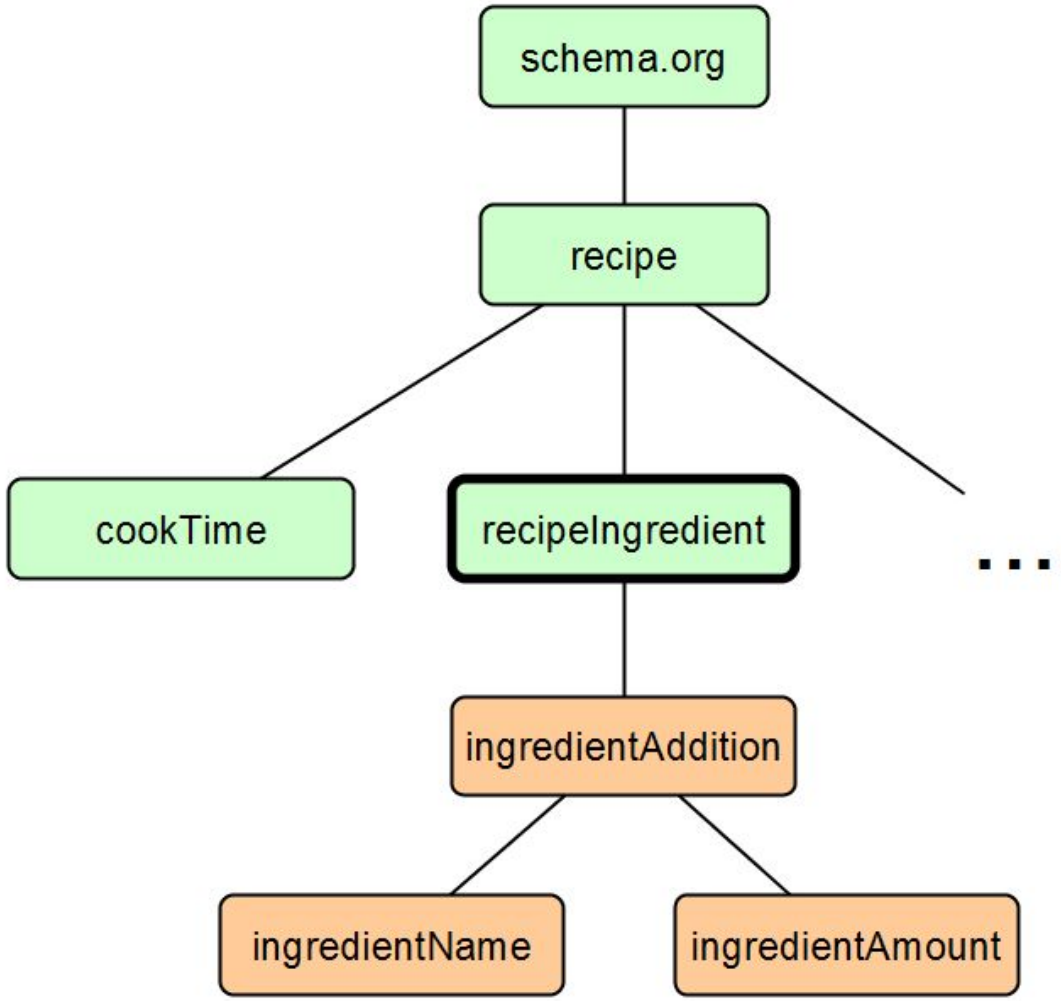
- Biggest and main supplier of data: EDAMAM
- API returns json file
- Is the richest in regards to information

Field	Type	Description
`uri`	**string**	Ontology identifier
`label`	**string**	Recipe title
`image`	**string**	Image URL
`source`	**string**	Source site identifier
`url`	**string**	Original recipe URL
`yield`	**integer**	Number of servings
`calories`	**float**	Total energy, kcal
`totalWeight`	**float**	Total weight, g
`ingredients`	**Ingredient[]**	Ingredients list
`totalNutrients`	**NutrientInfo[]**	Total nutrients for the entire recipe
`totalDaily`	**NutrientInfo[]**	% daily value for the entire recipe
`dietLabels`	**enum[]**	Diet labels: “balanced”, “high-protein”, “high-fiber”, “low-fat”, “low-carb”, “low-sodium” (labels are per serving)
`healthLabels`	**enum[]**	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
<http://schema.org/IngredientAddition>	32465
<http://schema.org/CreativeWork>	4433
<http://schema.org/NutritionInformation>	4433
<http://schema.org/Person>	4433
<http://schema.org/Recipe>	4433
<http://schema.org/Ingredient>	32465
<http://schema.org/QuantitativeValue>	36898
<http://schema.org/ImageObject>	4433

Knowledge Graph Statistics

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

Knowledge Graph Statistics

- Properties in Top5 Classes

Property	Class
<http://schema.org/unitText>	<http://schema.org/QuantitativeValue>
<http://schema.org/value>	<http://schema.org/QuantitativeValue>
<http://schema.org/ingridientFullName>	<http://schema.org/Ingredient>
<http://schema.org/name>	<http://schema.org/Ingredient>
<http://schema.org/ingredientName>	<http://schema.org/IngredientAddition>
<http://schema.org/ingridientAmount>	<http://schema.org/IngredientAddition>
<http://schema.org/url>	<http://schema.org/CreativeWork>
<http://schema.org/contentUrl>	<http://schema.org/ImageObject>
<http://schema.org/caption>	<http://schema.org/ImageObject>

Knowledge Graph Statistics

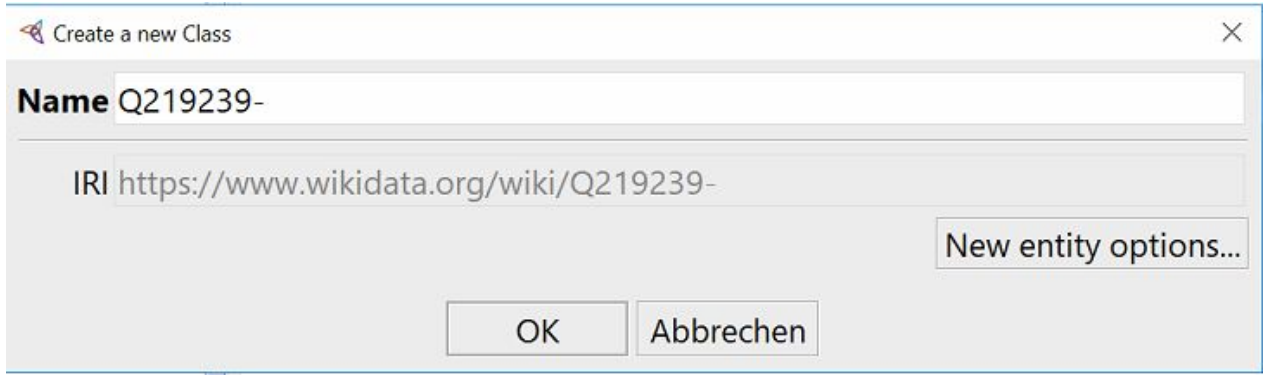
- Data Alignment

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

LOD Linking

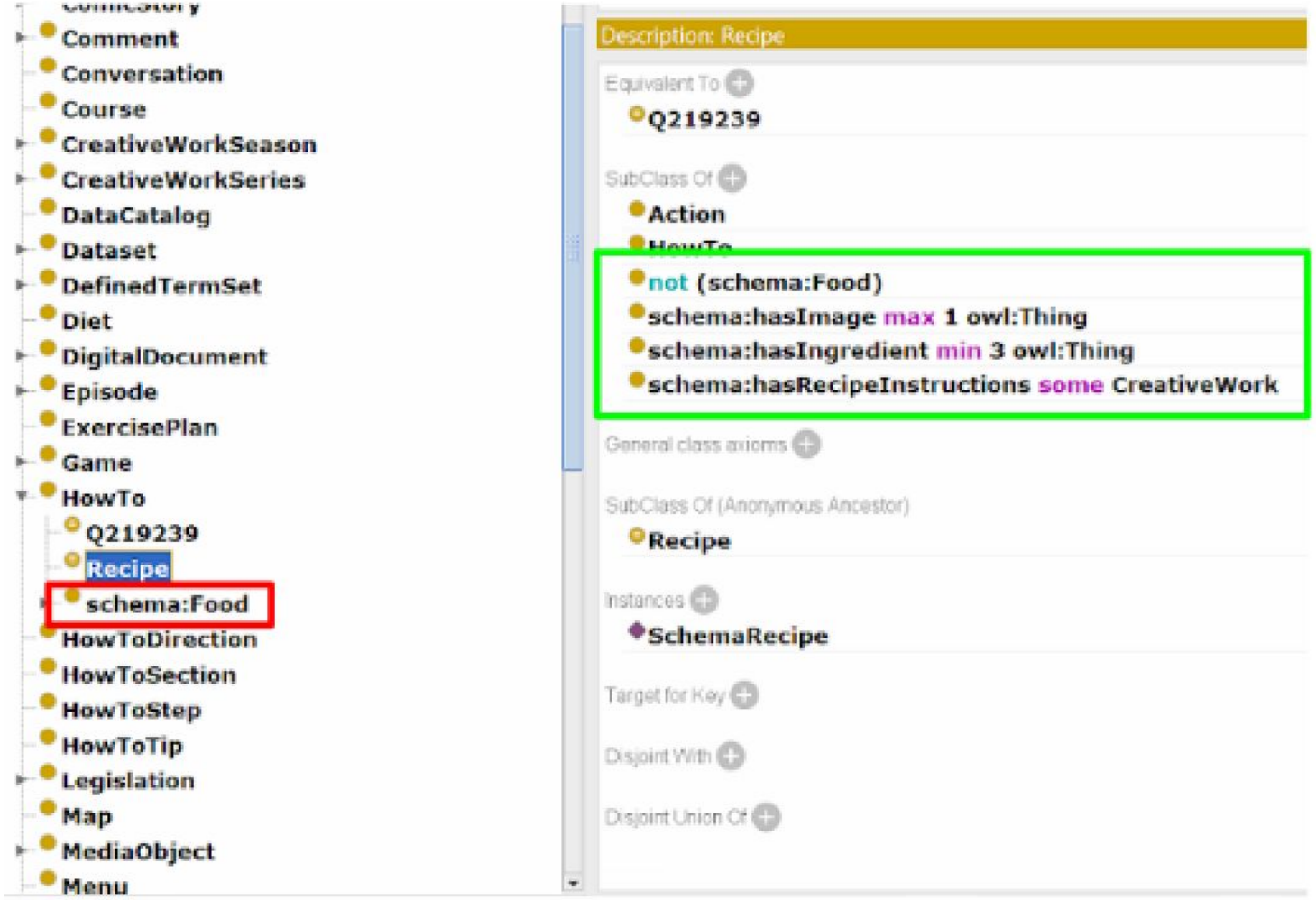
- creation of an entry
  - specify the wikidata-IRI

- name the entity the same as the Q identifier from wikidata (Recipe = Q219239)



LOD Linking

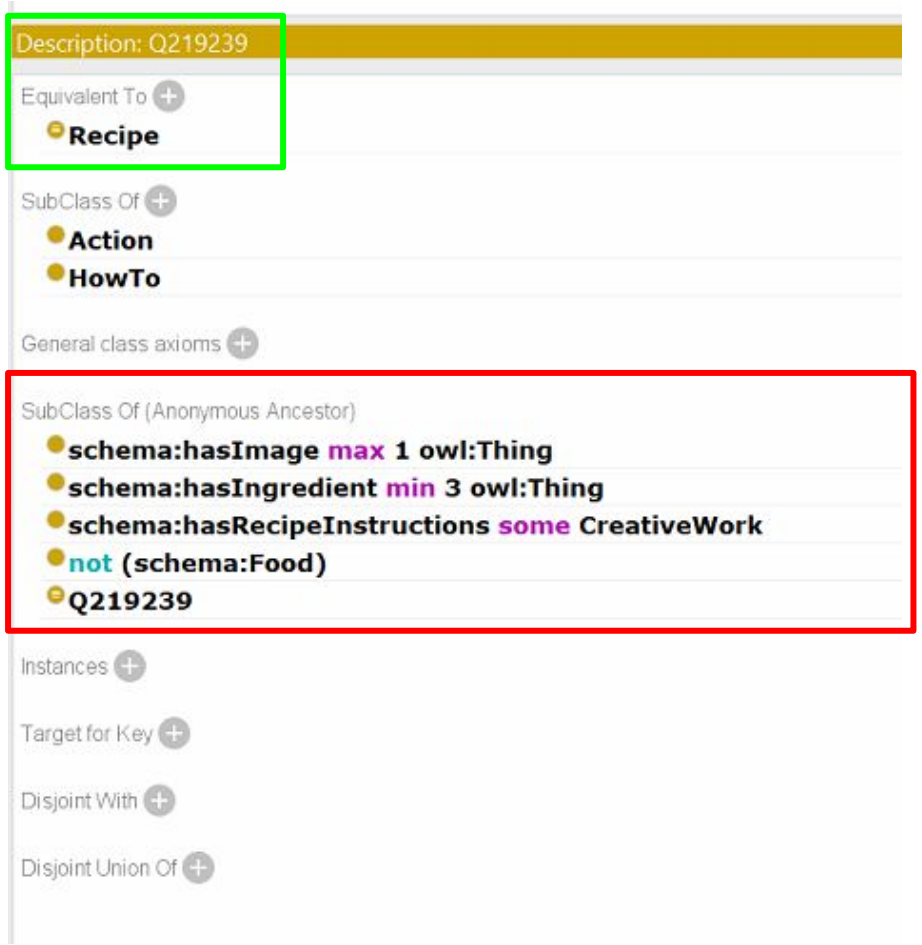
- imported the schema-vocabulary to Protege
- extended it with our own vocab-extensions (schema:Food)
- stated OWL axioms for the Recipe class



LOD Linking

- wikidata entity of recipe is set to equal with our recipe from schema

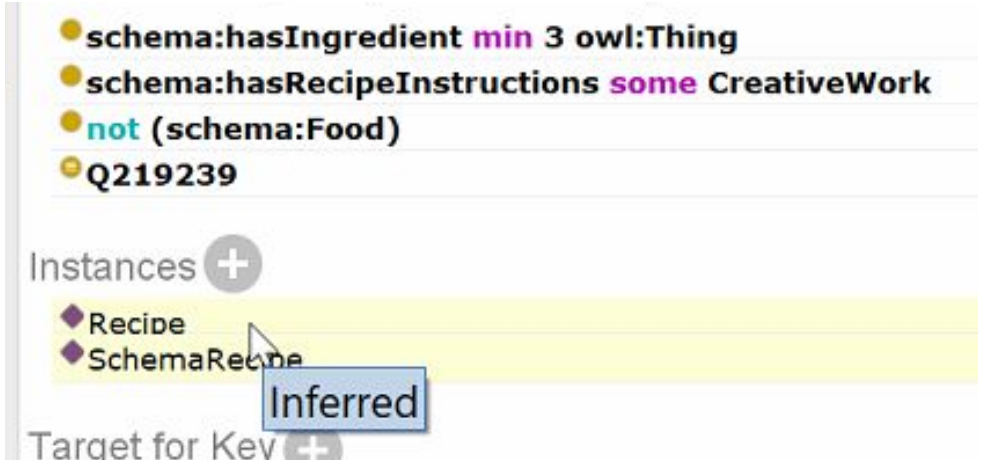
- subclasses are automatically inferred by the equivalence to the schema recipe entry



LOD Linking - Reasoner Results

- reasoner HermiT 1.3.8.413

- explanation of the reasoner why something is inferred



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

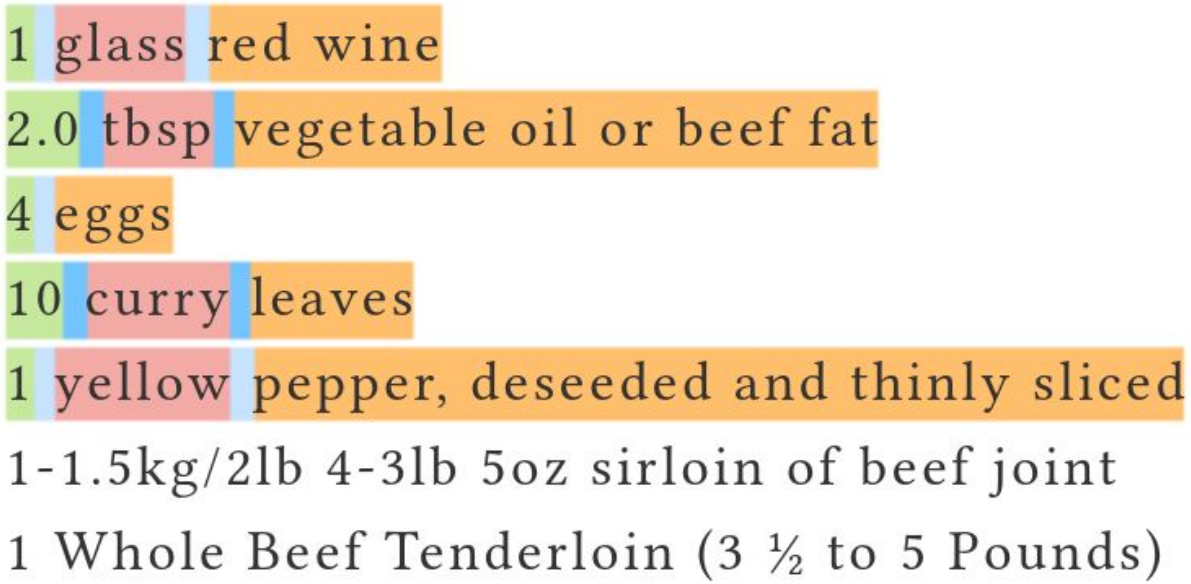
```
sh:property [
   sh:path schema:recipeIngredient ;
   sh:class schema:IngredientAddition ;
   sh:maxCount 15 ;
   sh:minCount 2 ;
] .
```

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



Naive Regex Parser Examples

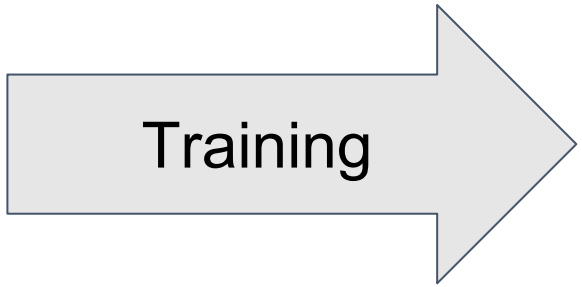
Information Extraction (NLP)

Use ML (NER) instead

```
1         QUANTITY
glass     UNIT
red       NAME
wine      NAME

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

...
```



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