

---

---

# Towards Food and Health Knowledge Graphs: From Language Models to Food Semantic Resources

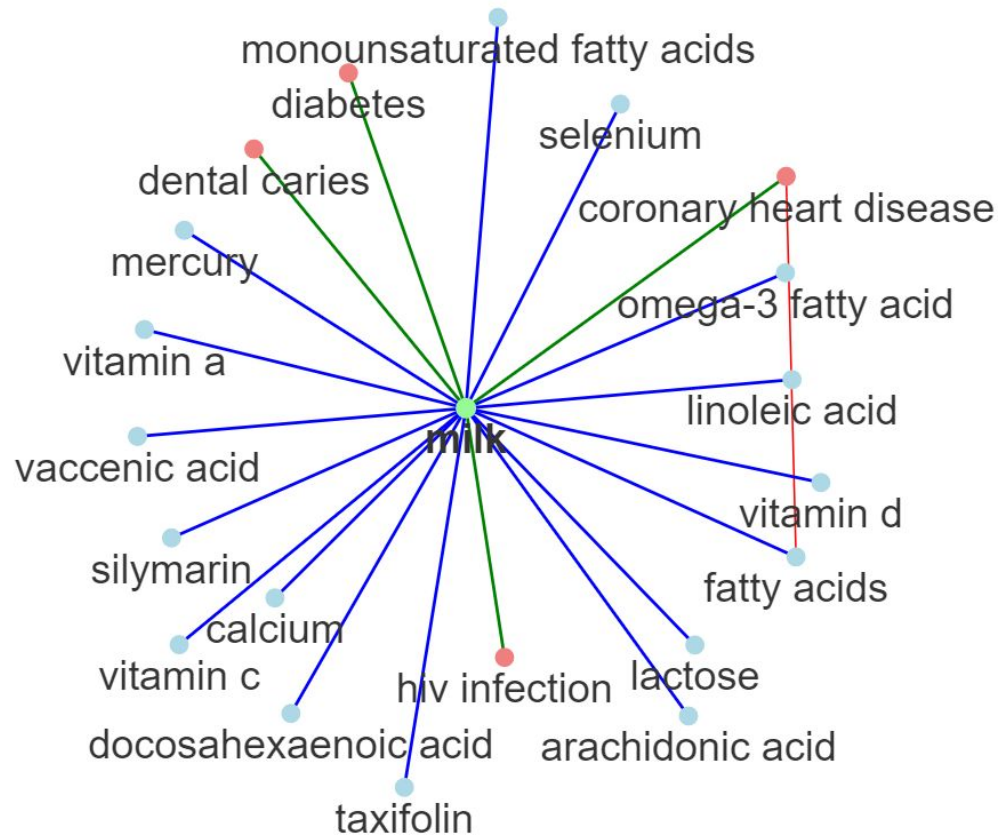
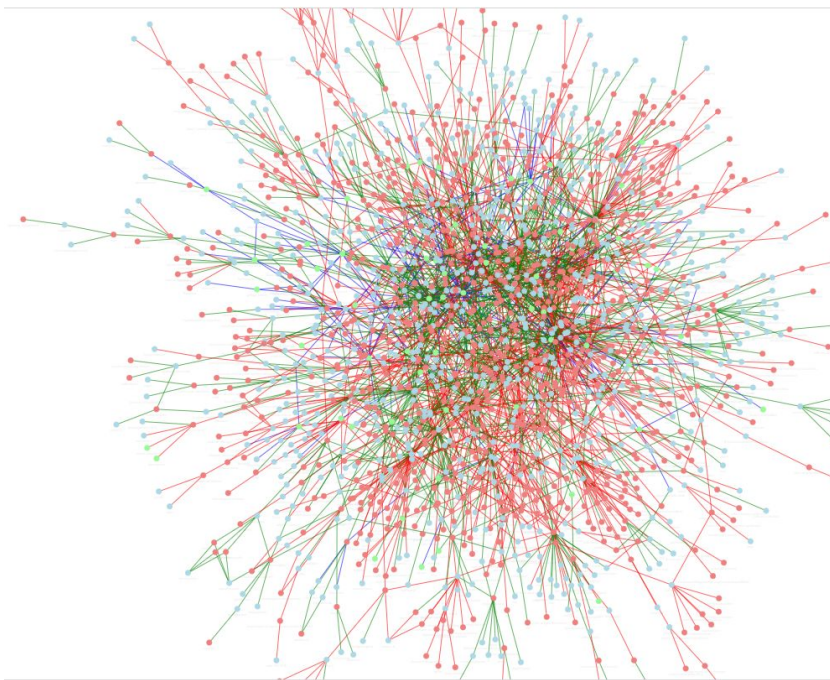
Tome Eftimov  
Computer Systems Department,  
Jožef Stefan Institute,  
Ljubljana, Slovenia

---

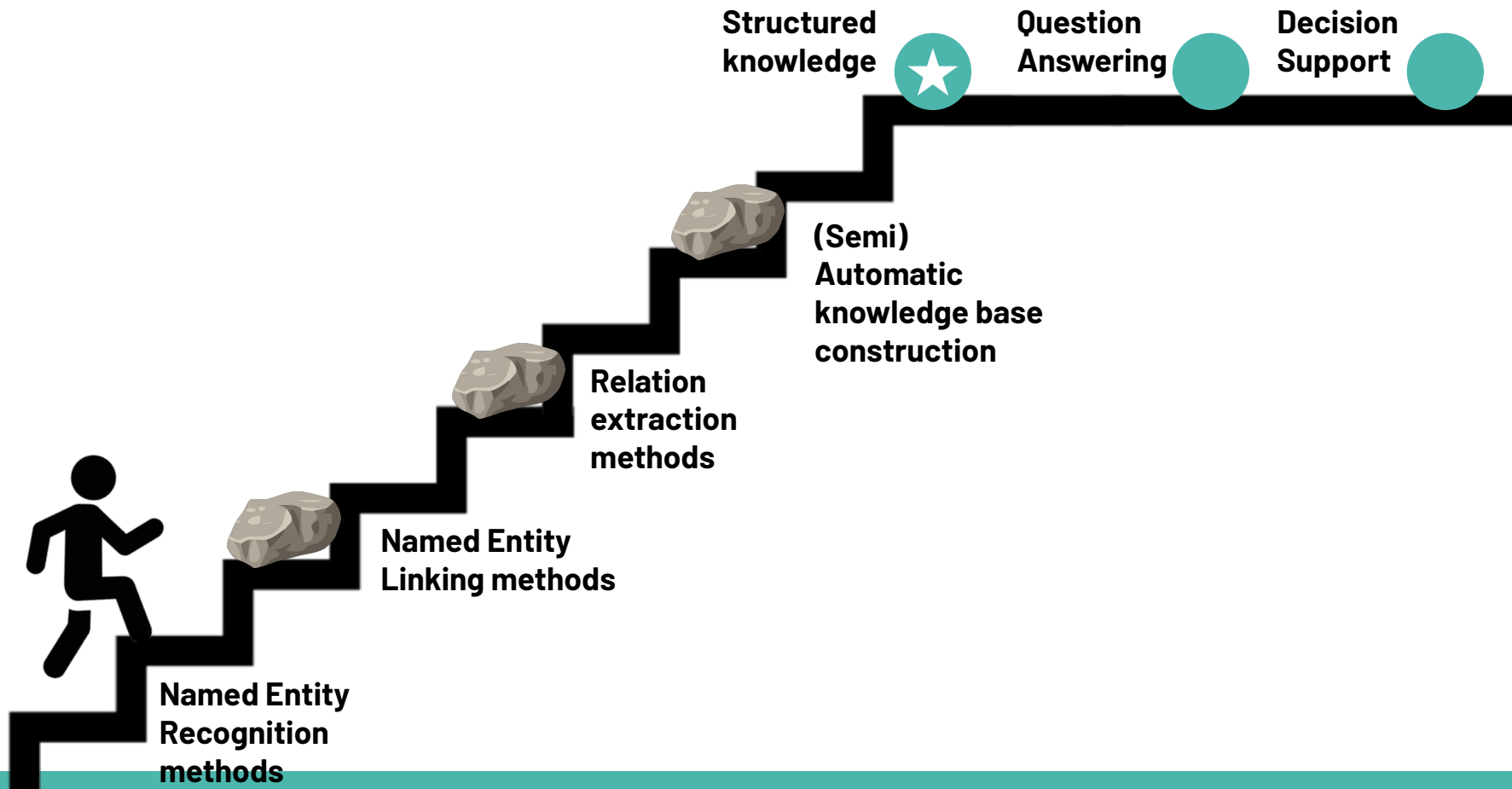
---

# Motivation

All extracted relations



# From Language Technologies to Decision Support



# Information Extraction (IE)

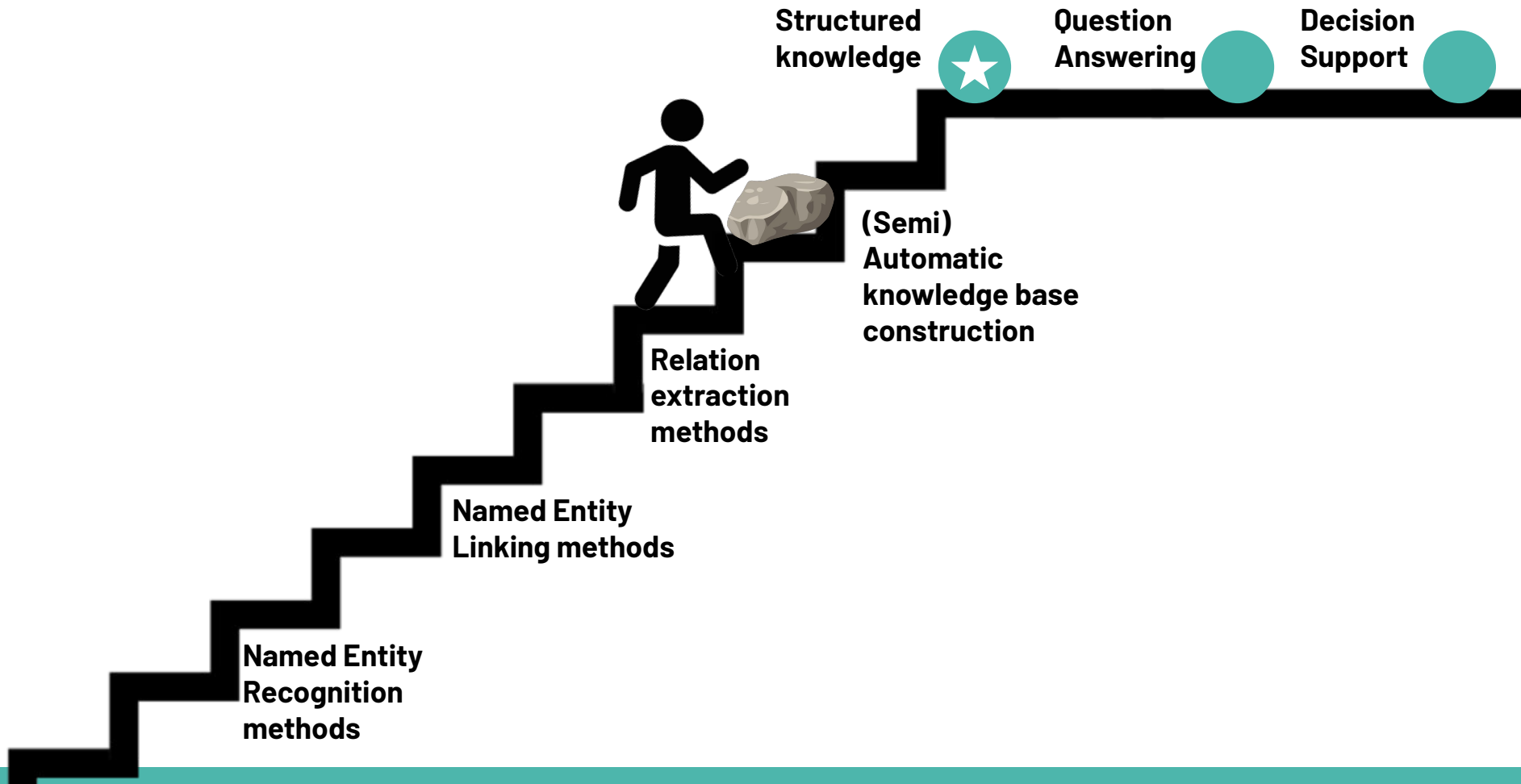
## Named-Entity Recognition (NER)

Excessive **salt** intake has been associated with a higher incidence of **heart disease**.

## Named-Entity Linking (NEL)

Excessive **salt** [**00002 (FOODB)**] intake has been associated with a higher incidence of **heart disease** [**0001 (UMLS)**].

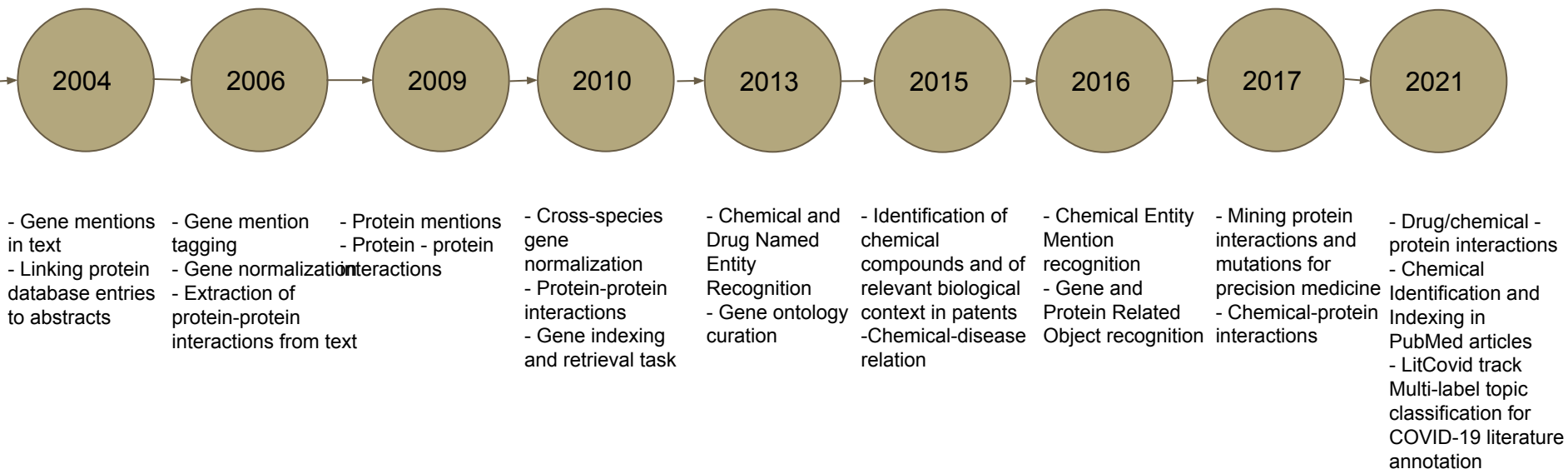
# State in the Biomedical Domain



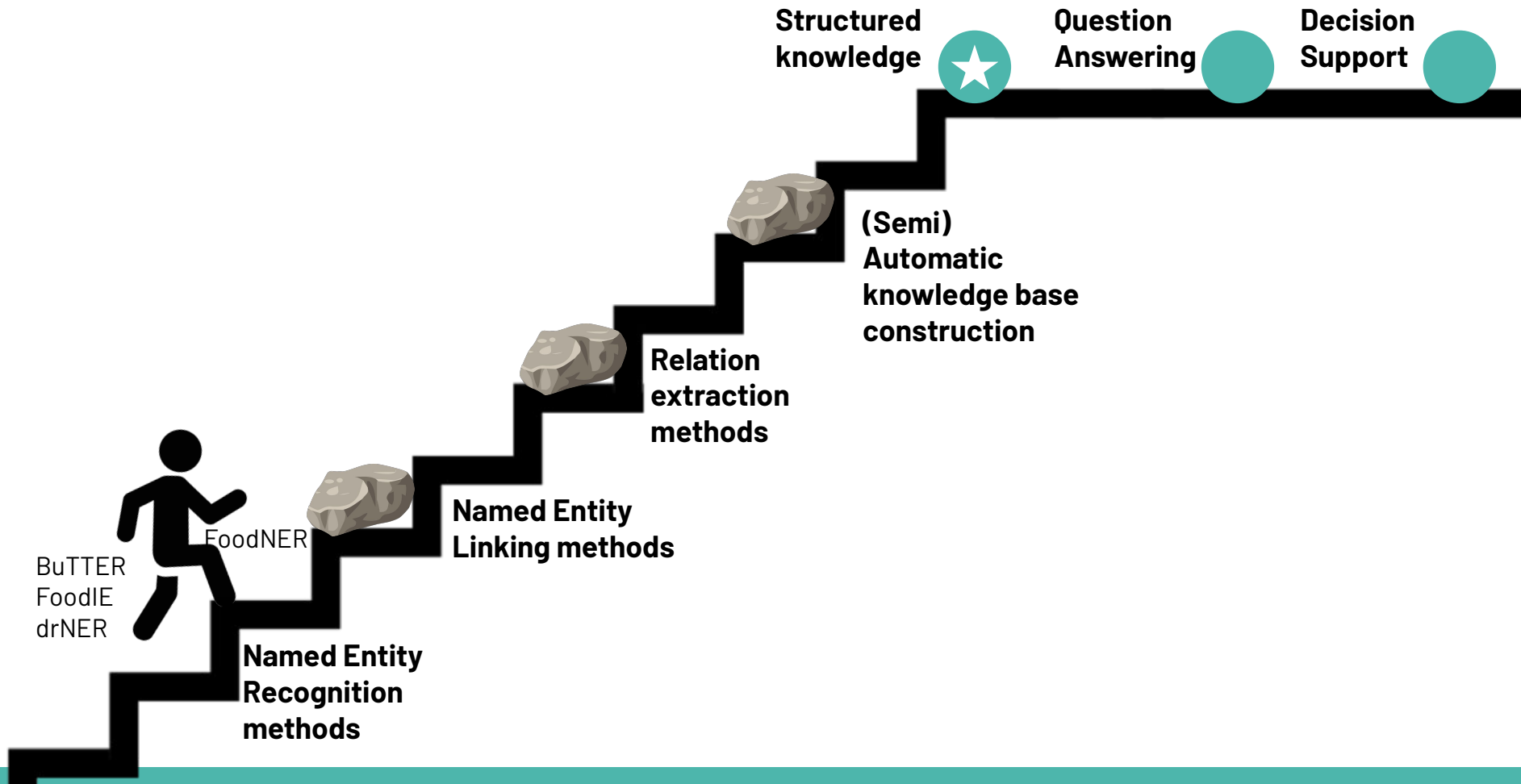
# Bio-IE During the Years

Several Bio-shared workshops

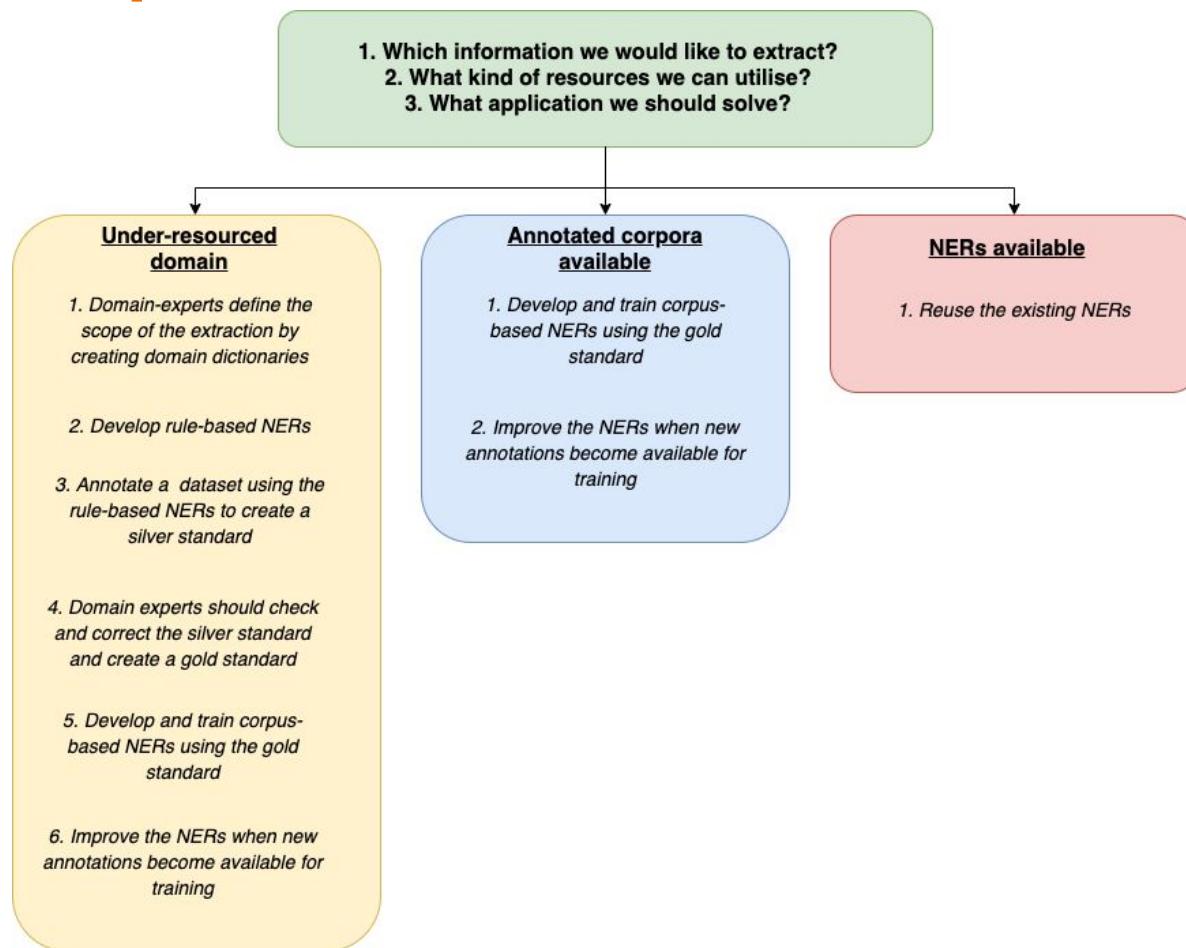
- BioNLP, BioCreative, ...



# State in the Food Domain



# CAFETERIA Pipeline





# CAFETERIA Methods

















- Rule-based methods:
  - spaCy + different food dictionaries
  - NCBO annotator with different food ontologies
- Corpus-based methods:
  - BuTTeR - bidirectional LSTM
    - ~ 94% macro averaged F1 score
  - FoodNER - fine-tuning BERT language model
    - ~ 94% macro averaged F1 score
    - ~ 73 - 78% macro averaged scores with regard to FoodOn, SNOMED-CT, and Hansard taxonomy

# Human-Computer Interaction Tool

## Recognized Entities for recipe 0recipe101

Preheat oven to 350 degrees F ( 175 degrees C ) . In a 12 inch skillet over medium heat , cook and stir the **garlic** and white parts of the **green onions** in **canola oil** until tender . Mix in **shredded chicken** , **salt** and **pepper** . Toss until well coated with **oil** . Stir in the **salsa** . Arrange **tortilla chips** on a large baking sheet . Spoon the **chicken mixture** over **tortilla chips** . Top with **Cheddar / Monterey Jack cheese blend** and **tomato** . Bake in the preheated oven 10 minutes , or until **cheese** has melted . Remove from heat and sprinkle with green onion tops before serving .

### Entity tags

Entity	Synonyms	Hansard Tags	Hansard Parent	Hansard Closest	FoodOn	SnomedCT	OF
garlic 	GARLIC	<div>AG.01.h.02.e [Onion/leek/garlic] </div>	AG.01.h	AG.01.h.02.e	<a href="#">obo:NCBITaxon_4682</a>	<a href="#">smct:735030001</a>	of:Garlic
green onions 		<div>AG.01.h.02.e [Onion/leek/garlic] </div>					
canola oil 	CANOLA OIL	<div>AG.01.f [Fat/oil] </div>	AG.01.f	AG.01.f	<a href="#">obo:FOODON_03310387</a> <a href="#">obo:FOODON_03302578</a>	<a href="#">smct:713648000</a> <a href="#">smct:226908006</a>	
shredded chicken 	SHREDDED CHICKEN CHICKEN	<div>AG.01.d.06 [Fowls] </div>	AG.01.d	AG.01.d.06	<a href="#">obo:NCBITaxon_9031</a> <a href="#">obo:FOODON_03411457</a>	<a href="#">smct:226955001</a>	of:Chicken
salt 	SALT	<div>AG.01.l.01 [Salt] </div>	AG.01.w	AG.01	<a href="#">obo:CHEBI_24866</a>		
pepper 		<div>AG.01.l.03 [Spice] </div>					
oil 	OIL	<div>AG.01.f [Fat/oil] </div>	AG.01.f	AG.01	<a href="#">obo:FOODON_03310387</a>		
salsa 	SALSA	<div>AG.01.l.04 [Sauce/dressing] </div>	AG.01.l	AG.01.l.04	<a href="#">obo:FOODON_03310086</a>		

# CAFETERIA Outcomes

- Annotated food consumption data:
  - 1000 recipes annotated with regard to Hansard, SNOMED-CT and FoodON
  - BioC format
- Annotated scientific abstracts :
  - 500 scientific abstracts annotated with regard to Hansard, SNOMED-CT and FoodON
  - BioC format

# An Example of Annotated Scientific Abstract

FoodViz Recipes Free text FoodNER annotation FoodNER resources Food Onto Map Index Food-Disease annotations Cafeteria annotations

Logout

c-10048971

c-10050267

c-10362132

c-10479224

c-10479227

c-10673913

c-10725162

c-10890036

c-10917925

c-10940346

c-10953669

c-10961155

c-11058886

c-11103296

c-11110871

c-11117618

c-11138444

c-1123503

c-11242456

c-11242460

c-11298008

c-11298013

c-11341048

c-11592684

c-11694649

c-11694651

c-11762742

c-11802218

c-11840174

c-11853496

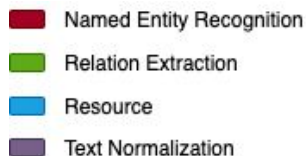
## Recognized Entities for recipe c-10050267

High intake of **red meat** or **processed meat** is associated with increased risk of colon cancer . In contrast , consumption of **white meat** ( **chicken** ) is not associated with risk and might even reduce the occurrence of colorectal cancer . We speculated that a diet containing **beef** or **bacon** would increase and a diet containing **chicken** would decrease colon carcinogenesis in rats . One hundred female Fischer 344 rats were given a single injection of azoxymethane ( 20 mg / kg i.p. ) , then randomized to 10 different AIN-76-based diets . Five diets were adjusted to 14 % fat and 23 % protein and five other diets to 28 % fat and 40 % protein . Fat and protein were supplied by 1 ) **lard** and casein , 2 ) **olive oil** and casein , 3 ) **beef** , 4 ) **chicken with skin** , and 5 ) **bacon** . Meat diets contained 30 % or 60 % freeze-dried fried **meat** . The diets were given ad libitum for 100 days , then colon tumor promotion was assessed by the multiplicity of aberrant crypt foci [ number of crypts per aberrant crypt focus ( ACF ) ] . The ACF multiplicity was nearly the same in all groups , except bacon-fed rats , with no effect of fat and protein level or source (  $p = 0.7$  between 8 groups by analysis of variance ) . In contrast , compared with lard - and casein-fed controls , the ACF multiplicity was reduced by 12 % in rats fed a diet with 30 % **bacon** and by 20 % in rats fed a diet with 60 % **bacon** (  $p < 0.001$  ) . The **water intake** was higher in bacon-fed rats than in controls (  $p < 0.0001$  ) . The concentrations of iron and bile acids in fecal water and total fatty acids in feces changed with diet , but there was no correlation between these concentrations and the ACF multiplicity . Thus the hypothesis that colonic iron , bile acids , or total fatty acids can promote colon tumors is not supported by this study . The results suggest that , in rats , **beef** does not promote the growth of ACF and **chicken** does not protect against colon carcinogenesis . A bacon-based diet appears to protect against carcinogenesis , perhaps because **bacon** contains 5 % NaCl and increased the rats' water intake .

### Entity tags

Entity	Synonyms	Hansard Tags	Hansard Parent	Hansard Closest	FoodOn	SnomedCT	OF
red meat <span>x</span>		[AG.01.d] Animals for food <span>v</span>	Animals for food	Animals for food			
processed meat <span>x</span>		[AG.01.d] Animals for food <span>v</span>	Animals for food	Animals for food			
white meat <span>x</span>		[AG.01.d.06] Fowls <span>v</span>	Animals for food	Fowls			
chicken <span>x</span>	CHICKEN	[AG.01.d.06] Fowls <span>v</span>	Food	Genus Gallus (domestic fowl)	Gallus gallus chicken	Chicken - meat	of:Chicken

# References



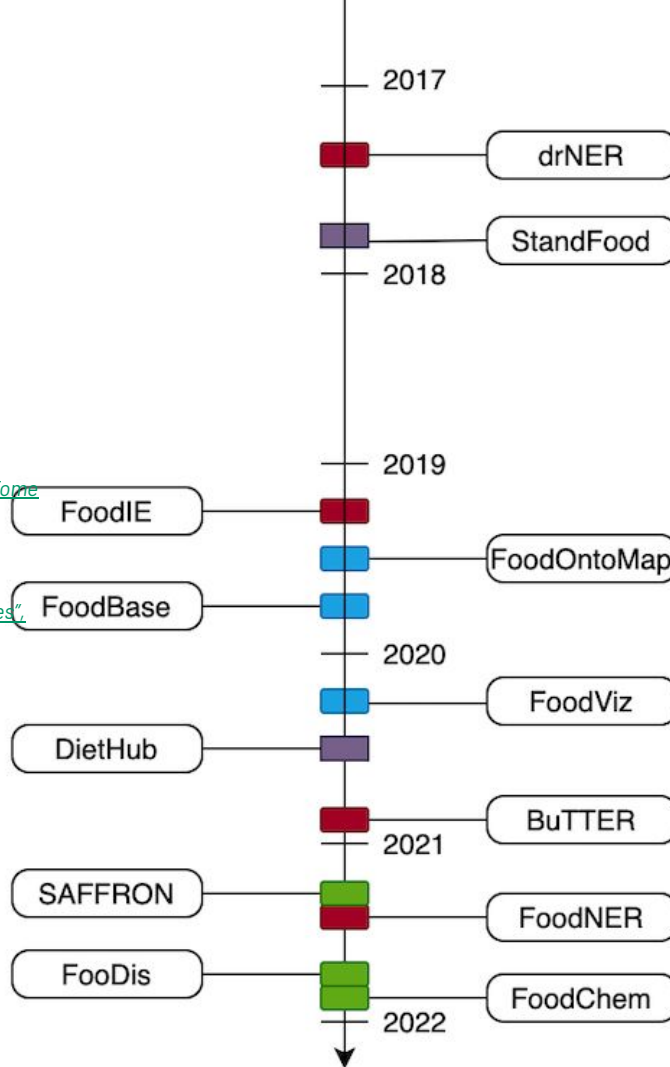
[Gorjan Popovski, Stefan Kochev, Barbara Koroušić Seljak, Tome Eftimov. "FoodIE: A Rule-based Named-entity Recognition Method for Food Information Extraction", ICPRAM 2019](#)

[Gorjan Popovski, Barbara Koroušić Seljak, Tome Eftimov. "FoodBase corpus: a new resource of annotated food entities", Database, 2019](#)

[Eftimov, T., Korošec, P., & Koroušić Seljak, B. \(2017\). StandFood: standardization of foods using a semi-automatic system for classifying and describing foods according to FoodEx2. \*Nutrients\*, 9\(6\), 542.](#)

[Gjorgjina Cenikj, Tome Eftimov, Barbara Koroušić Seljak. "SAFFRON: tranSfer leARning For Food-Disease RelatiOn extraction", NAACL 2021](#)

[Gjorgjina Cenikj, Tome Eftimov, Barbara Koroušić Seljak. "FoodDis: A food-disease relation mining pipeline", \*Expert Systems with Applications\*, 2021, In review](#)



[Tome Eftimov, Barbara Koroušić Seljak, Peter Korošec. "A rule-based named-entity recognition method for knowledge extraction of evidence-based dietary recommendations", \*PLoS One\*, 2017](#)

[Eftimov, T., Korošec, P., & Koroušić Seljak, B. \(2017\). StandFood: standardization of foods using a semi-automatic system for classifying and describing foods according to FoodEx2. \*Nutrients\*, 9\(6\), 542.](#)

[Gorjan Popovski, Barbara Koroušić Seljak, Tome Eftimov. "FoodOntoMap: Linking Food Concepts across Different Food Ontologies", \*KEOD\*, 2019](#)

[Riste Stojanov, Gorjan Popovski, Nasi Jofce, Dimitar Trajanov, Barbara Koroušić Seljak, Tome Eftimov. FoodViz: Visualization of Food Entities Linked Across Different Standards, \*LNCS\*, 2020](#)

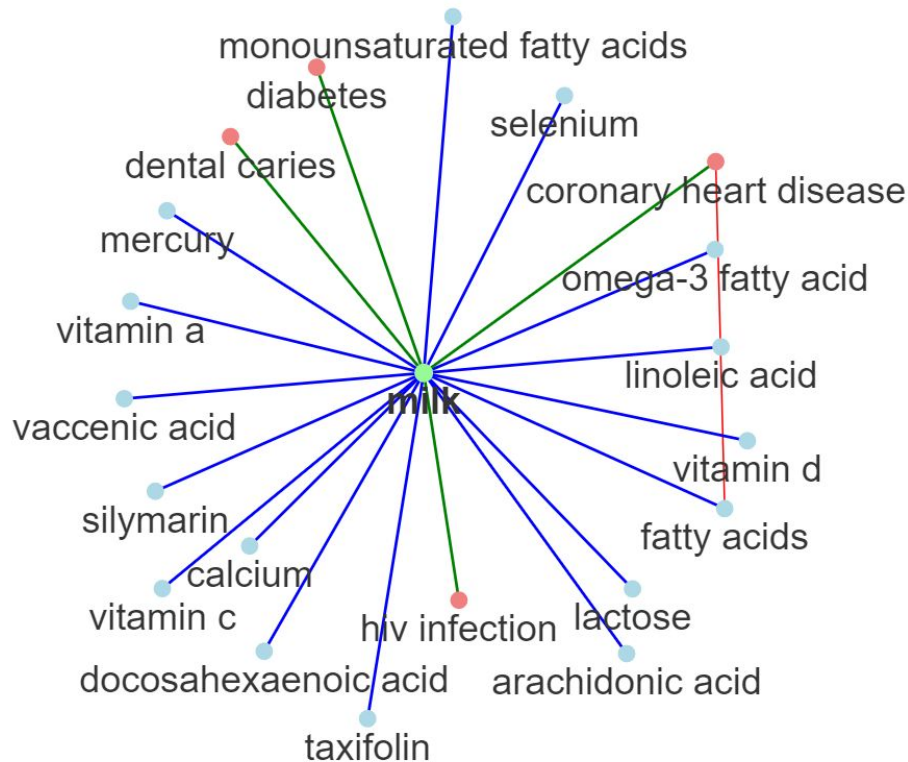
[Gjorgjina Cenikj, Gorjan Popovski, Riste Stojanov, Barbara Koroušić Seljak, Tome Eftimov. "BuTTER: Bidirectional LSTM for Food Named-Entity Recognition", \*BFNDMA 2020\*](#)

[Riste Stojanov, Gorjan Popovski, Gjorgjina Cenikj, Barbara Koroušić Seljak, Tome Eftimov. "A Fine-Tuned Bidirectional Encoder Representations From Transformers Model for Food Named-Entity Recognition: Algorithm Development and Validation", \*JMIR\* 2021](#)

[Cenikj, G., Seljak, B. K., & Eftimov, T. \(2021, December\). FoodChem: A food-chemical relation extraction model. In \*2021 IEEE Symposium Series on Computational Intelligence \(SSCI\)\* \(pp. 1-8\). \*IEEE\*.](#)

# Towards Knowledge Graphs

All extracted relations





# Team



Gjorgjina Cenikj  
Master Student  
JSI



Gordana Ispirova  
PhD Candidate  
JSI



Matevž Ogrinc  
Master Student  
JSI



Eva Valenčič  
PhD Candidate  
JSI



Asst. Prof. Riste Stojanov, PhD  
FINKI



Prof. Peter Korošec, PhD  
Senior Researcher  
JSI



Asst. Prof. Tome Eftimov, PhD  
Senior Researcher  
JSI



Prof. Barbara Koroušič Seljak, PhD  
Senior Researcher  
JSI