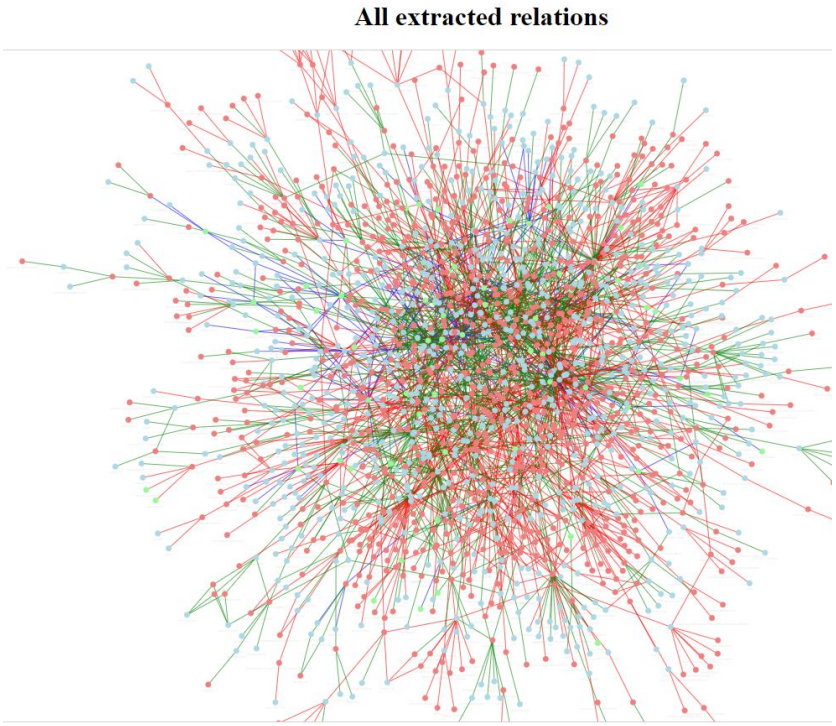
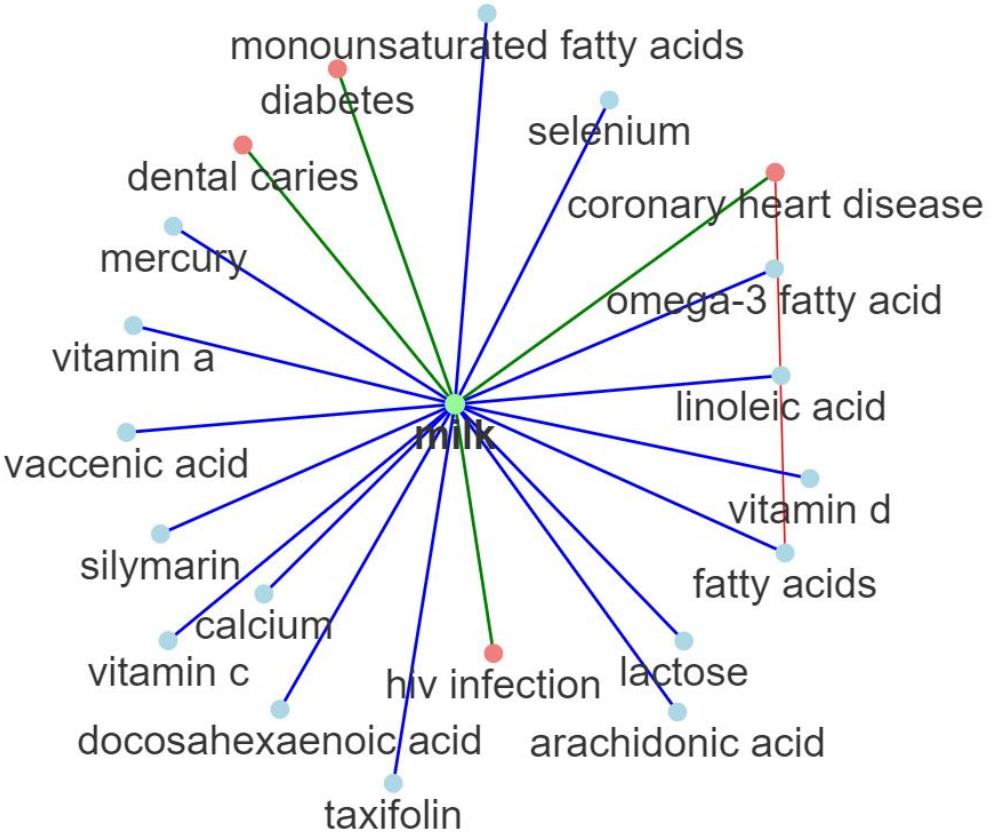
Towards Food and Health Knowledge Graphs: From Language Models to Food Semantic Resources

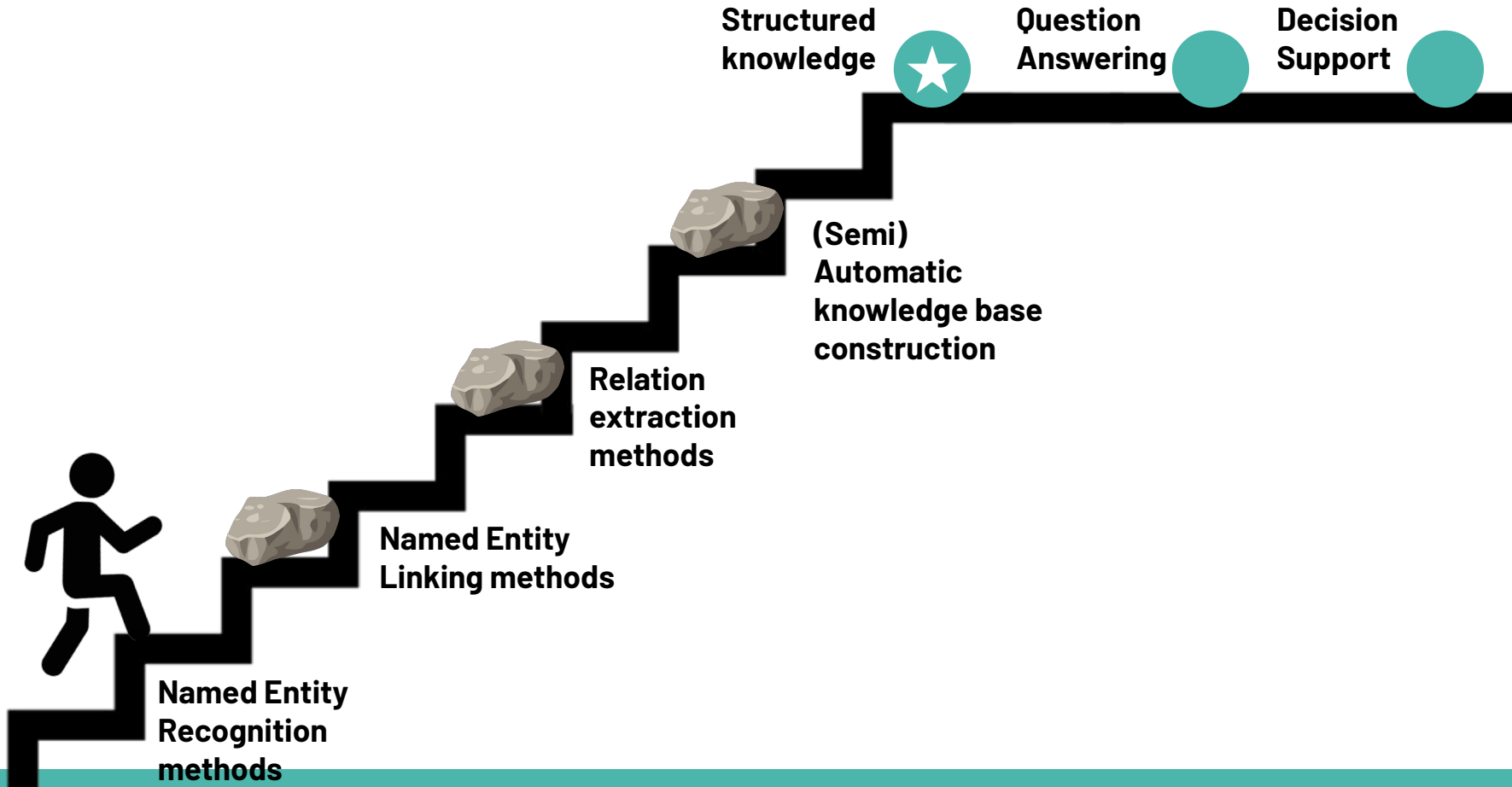
Tome Eftimov
Computer Systems Department,
Jožef Stefan Institute,
Ljubljana, Slovenia

Motivation


All extracted relations


monounsaturated fatty acids
diabetes
selenium
dental caries
coronary heart disease
mercury
omega-3 fatty acid
vitamin a
linoleic acid
milk
vaccenic acid
vitamin d
silymarin
fatty acids
calcium
vitamin c
hiv infection
lactose
docosahexaenoic acid
arachidonic acid
taxifolin

From Language Technologies to Decision Support



Information Extraction (IE)

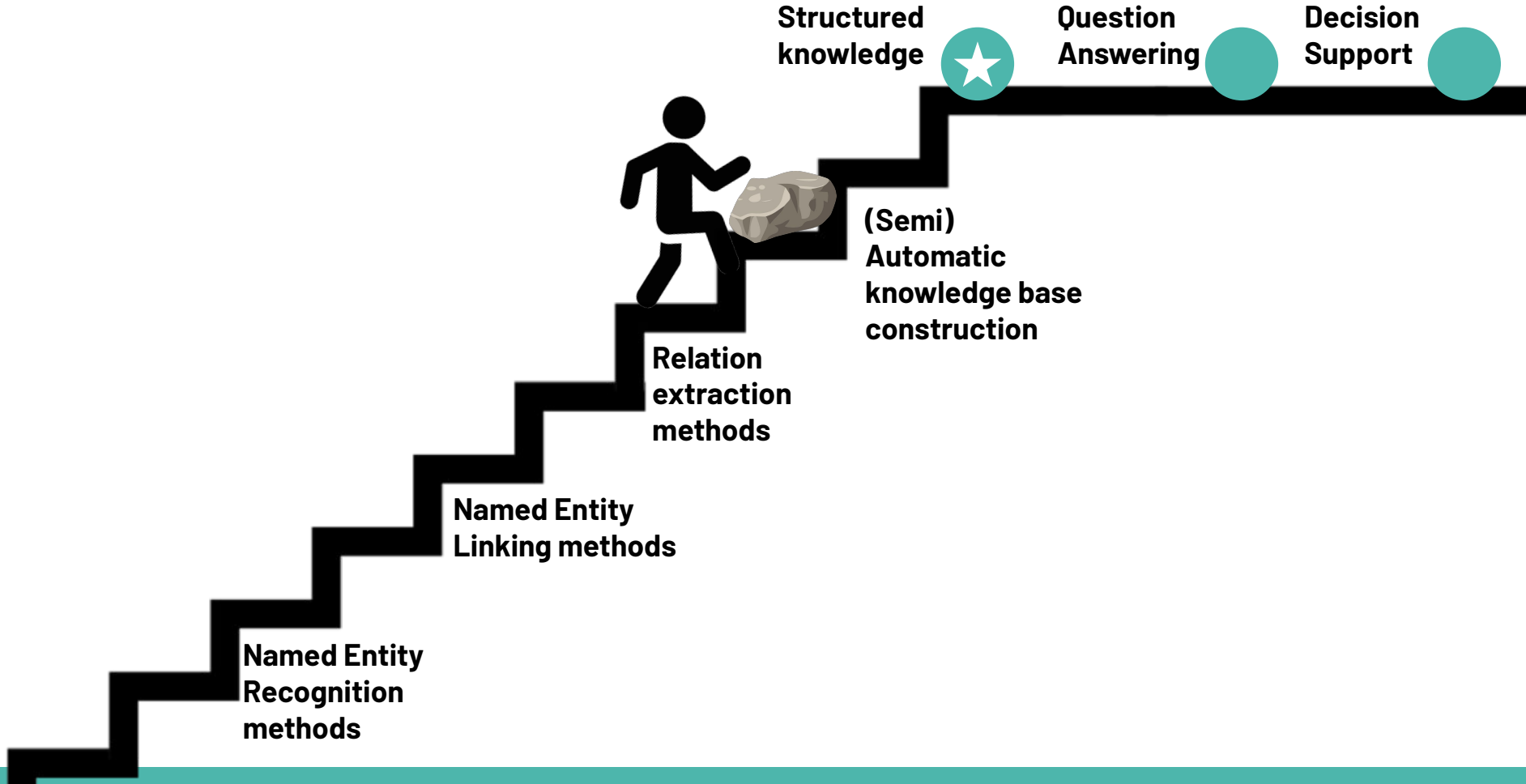
Named-Entity Recognition (NER)

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)]**.

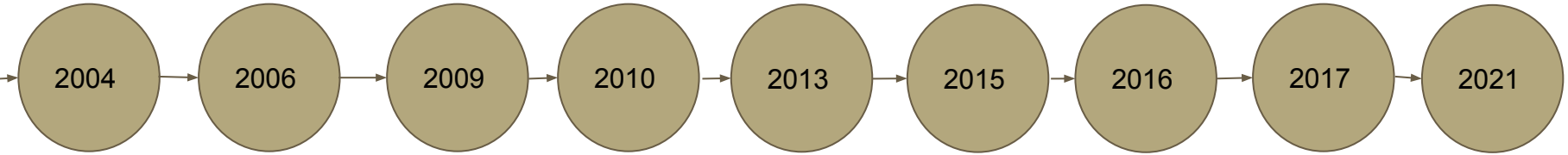
State in the Biomedical Domain



Bio-IE During the Years

Several Bio-shared workshops

- BioNLP, BioCreative, ...



- Gene mentions in text
- Linking protein database entries to abstracts

- Gene mention tagging
- Gene normalization
- Extraction of protein-protein interactions from text

- Protein mentions
- Protein - protein interactions

- Cross-species gene normalization
- Protein-protein interactions
- Gene indexing and retrieval task

- Chemical and Drug Named Entity Recognition
- Gene ontology curation

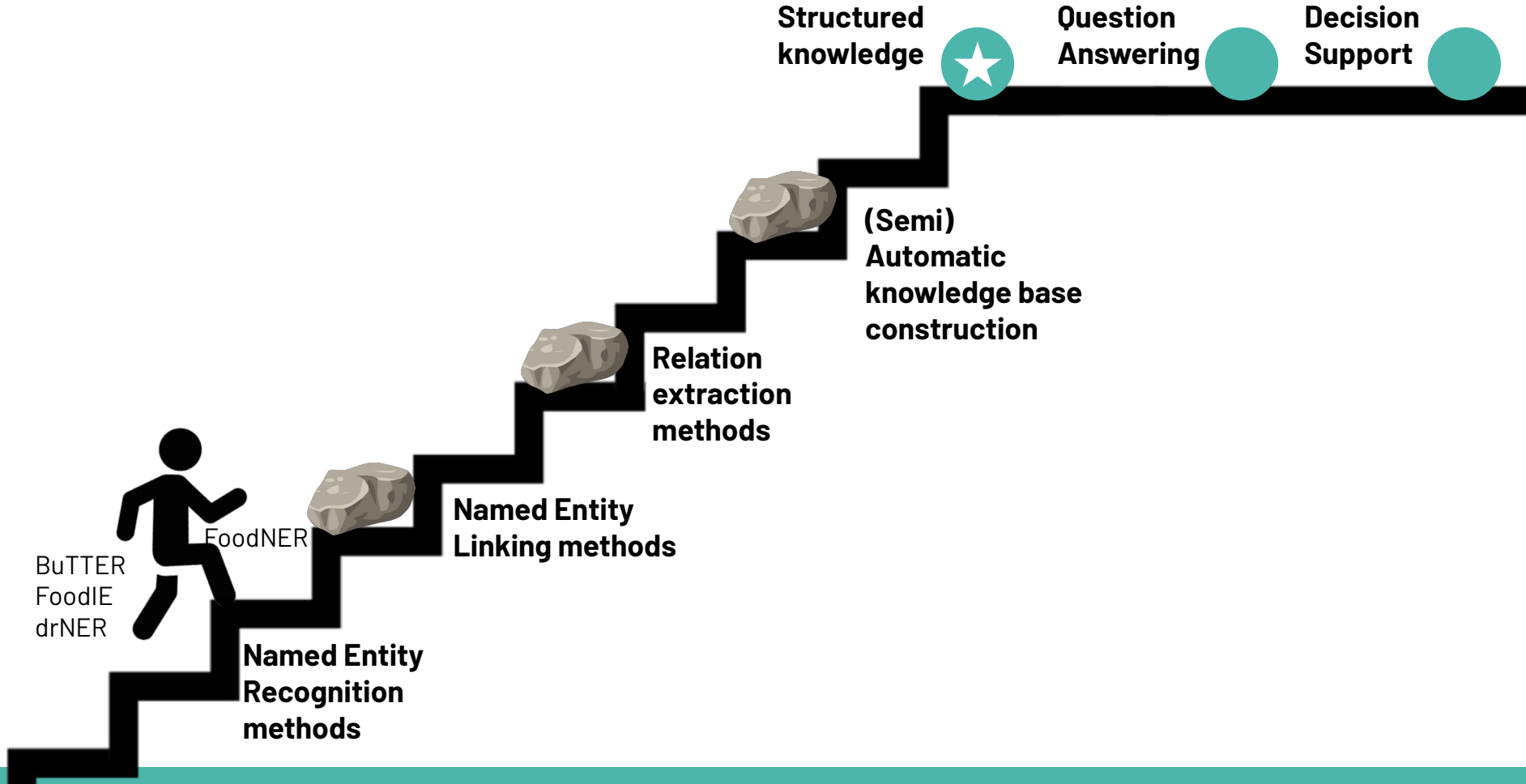
- Identification of chemical compounds and of relevant biological context in patents
-Chemical-disease relation

- Chemical Entity Mention recognition
- Gene and Protein Related Object recognition

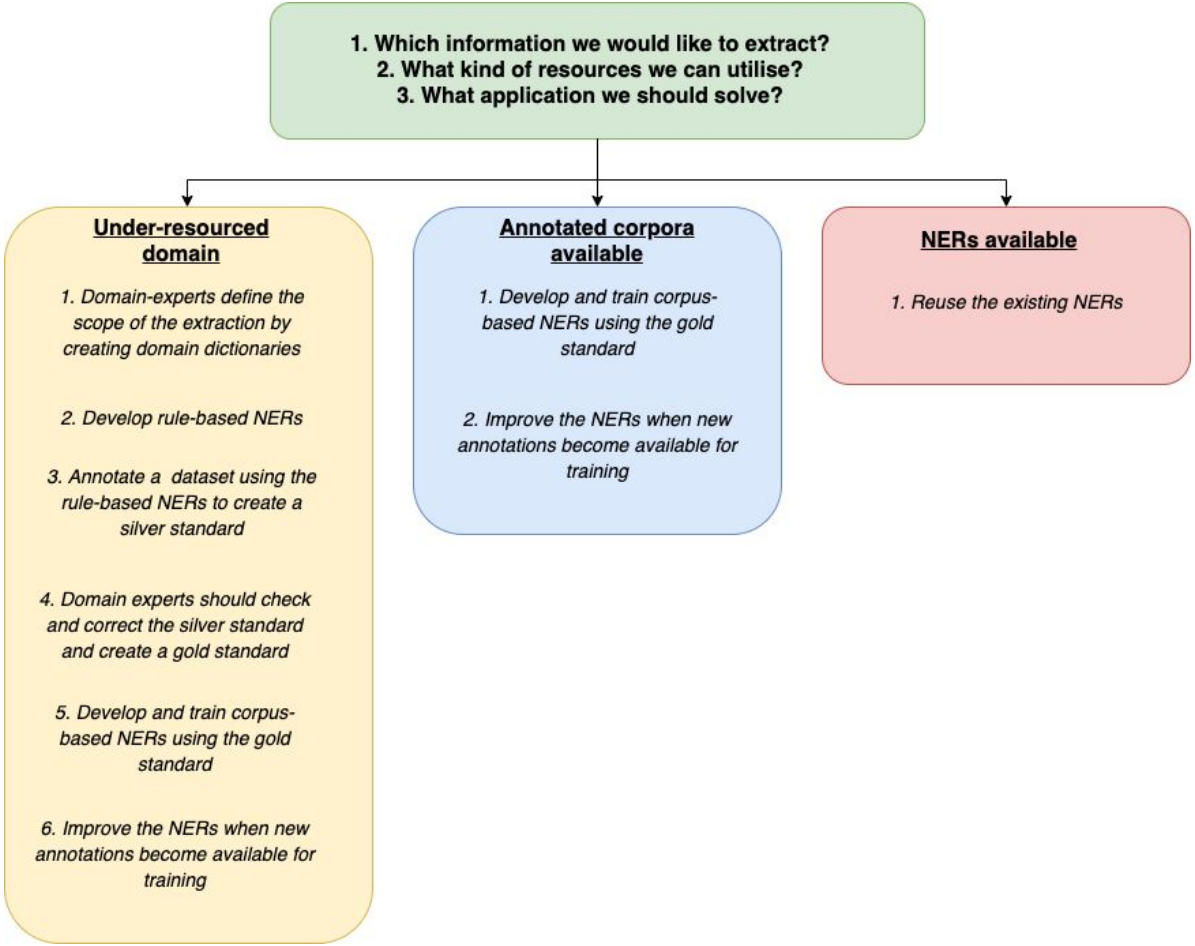
- Mining protein interactions and mutations for precision medicine
- Chemical-protein interactions

- Drug/chemical - protein interactions
- Chemical Identification and Indexing in PubMed articles
- LitCovid track Multi-label topic classification for COVID-19 literature annotation

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

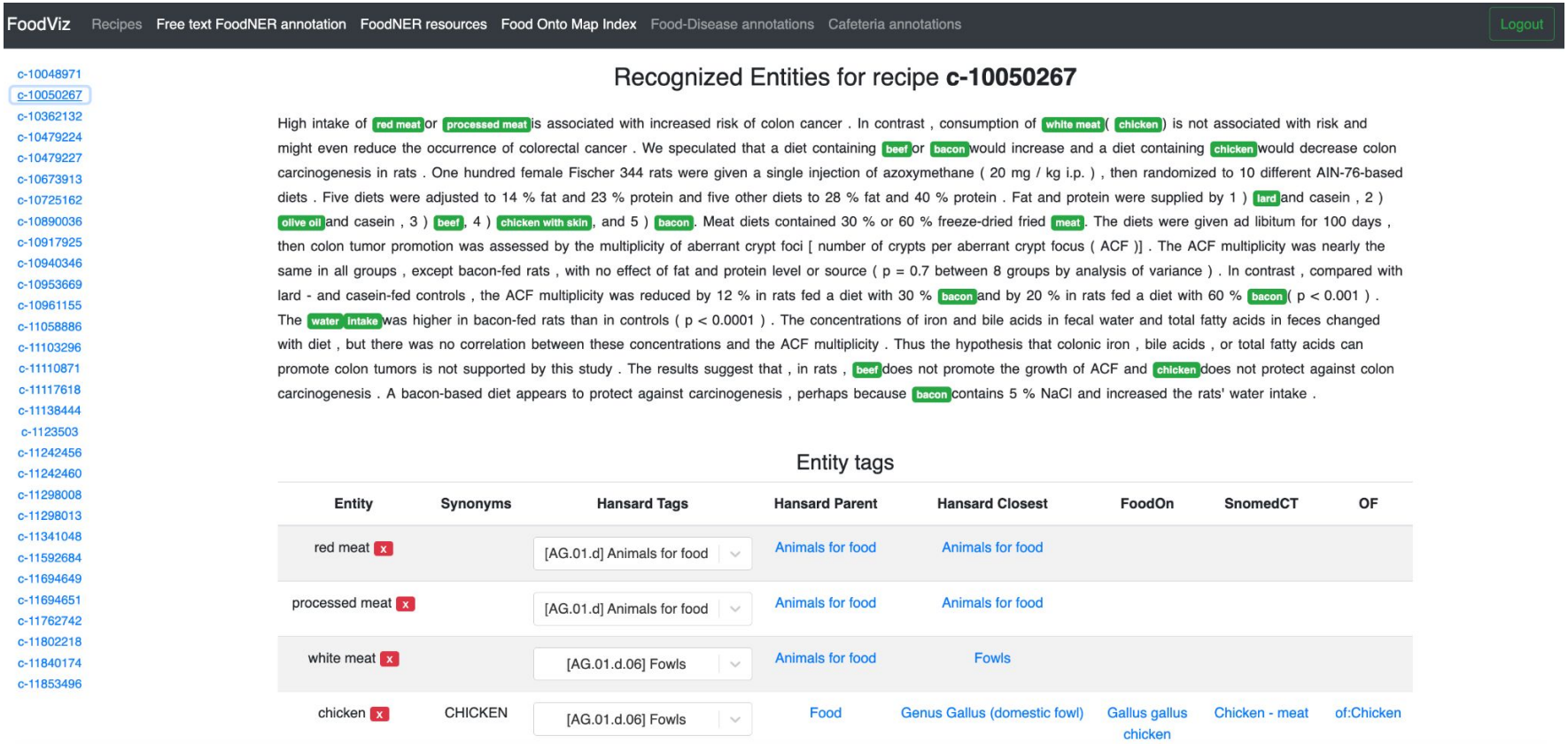
Entity	Synonyms	Hansard Tags	Hansard Parent	Hansard Closest	FoodOn	SnomedCT	OF
garlic x	GARLIC	AG.01.h.02.e [Onion/leek/garlic]	AG.01.h	AG.01.h.02.e	obo:NCBITaxon_4682	smct:735030001	of:Garlic
green onions x		AG.01.h.02.e [Onion/leek/garlic]					
canola oil x	CANOLA OIL	AG.01.f [Fat/oil]	AG.01.f	AG.01.f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

CAFETERIA Outcomes

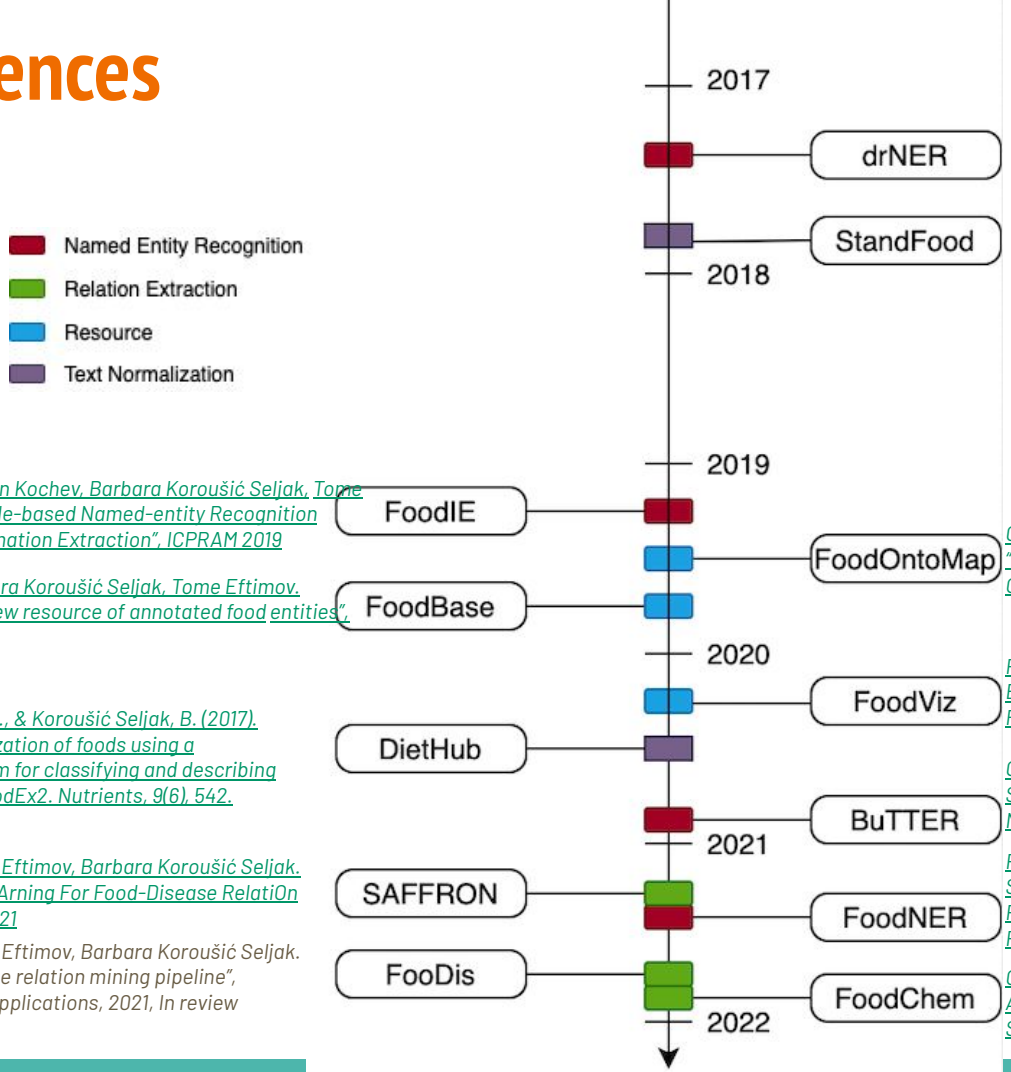
- Annotated food consumption data:
  - 1000 recipes annotated with regard to Hansard, SNOMED-CT and FoodON
  - BioC format

- Annotated scientific abstracts :
  - 500 scientific abstracts annotated with regard to Hansard, SNOMED-CT and FoodON
  - BioC format

An Example of Annotated Scientific Abstract



References



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, 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. "FoodOntoMap: Linking Food Concepts across Different Food Ontologies", KEOD, 2019

Gorjan Popovski, Barbara Koroušić Seljak, Tome Eftimov. "FoodBase corpus: a new resource of annotated food entities", Database, 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

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, Gorjan Popovski, Riste Stojanov, Barbara Koroušić Seljak, Tome Eftimov. "BuTTER: BidirecTional LSTM for Food Named-Entity Recognition", BFNDMA 2020

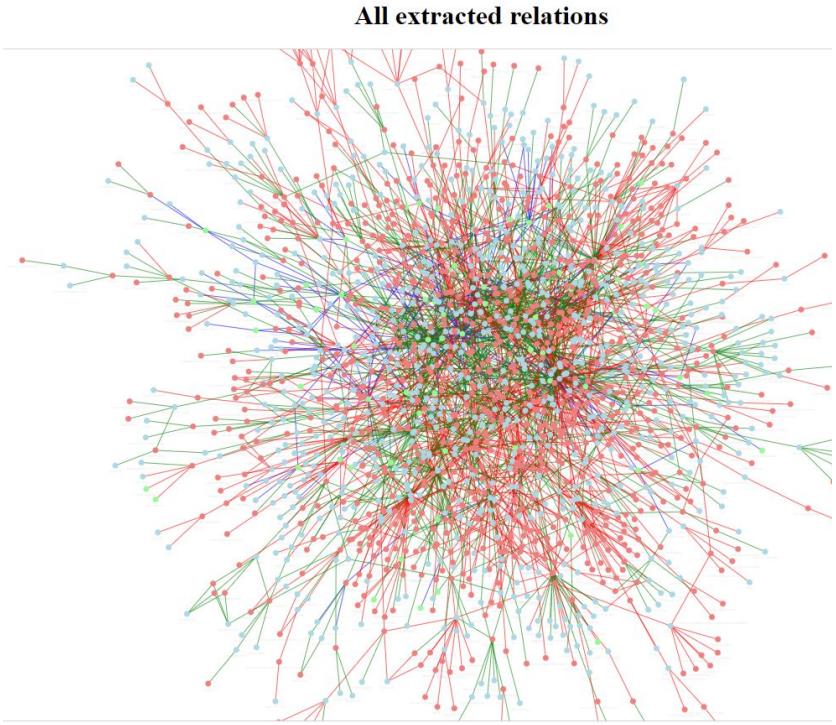
Gjorgjina Cenikj, Tome Eftimov, Barbara Koroušić Seljak. "SAFFRON: tranSfer leArning For Food-Disease RelatiOn extractioN", NAACL 2021

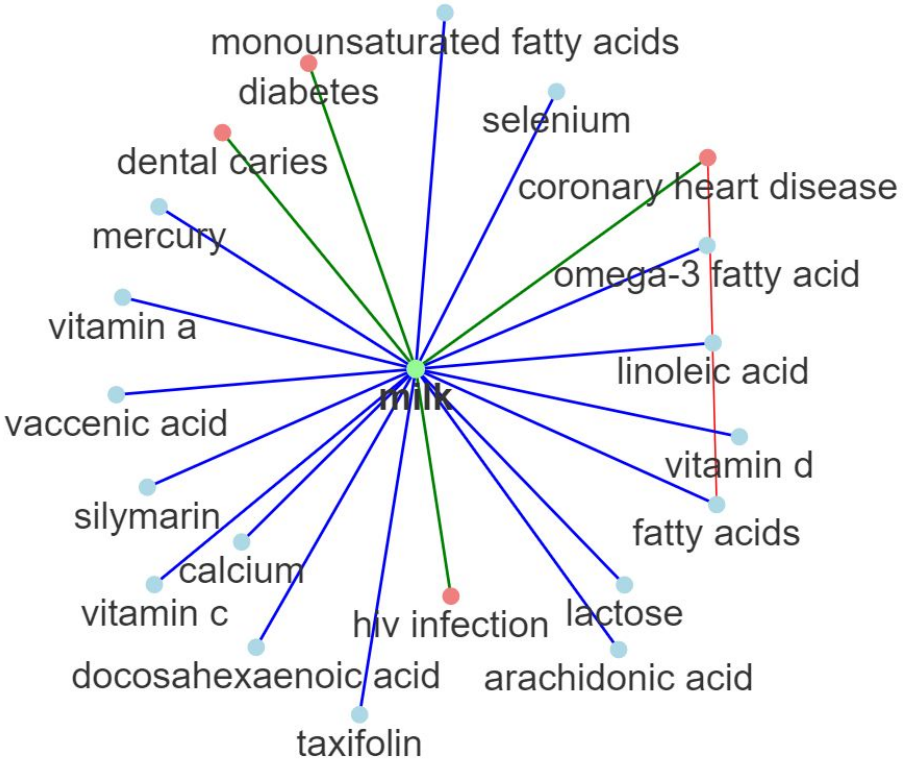
Riste Stojanov, Gorjan Popovski, Gjorgjina Cenikj, Barbara Koroušić Seljak, Tome Eftimov. "A Fine-Tuned Bidirectional Encoder Representations From Transformers Model for Food Named-Entity Recognition: Algorithm Development and Validation", JMIR 2021

Gjorgjina Cenikj, Tome Eftimov, Barbara Koroušić Seljak. "FooDis: A food-disease relation mining pipeline", Expert Systems with Applications, 2021, In review

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





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 Koroušić Seljak, PhD
Senior Researcher
JSI