

Creating a semantic food knowledge base with cooking recipes for a meal recommender system

Benjamin Gateau, Christoph Stahl, Olivier Pedretti

Senior R&T Associate,

Human Dynamics in Cognitive Environments Unit

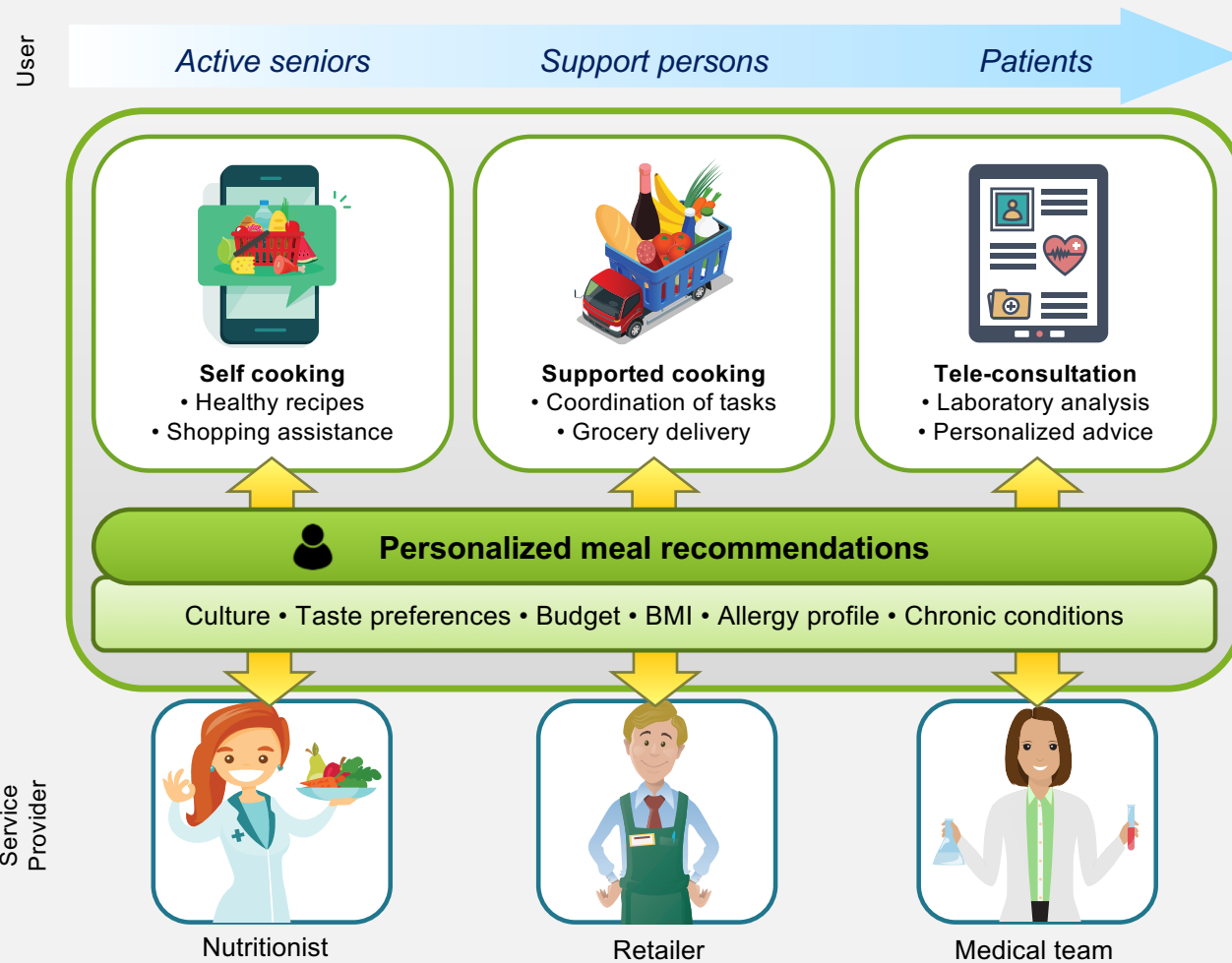
Luxembourg Institute of Science and Technology

Outline



- LIFANA – Lifelong Food and Nutrition Assistance
- Integration of recipes
- Alignment between national food information resources
 - Hybrid (text and conceptual) similarity measure

LIFANA Nutrition Solution



Goals

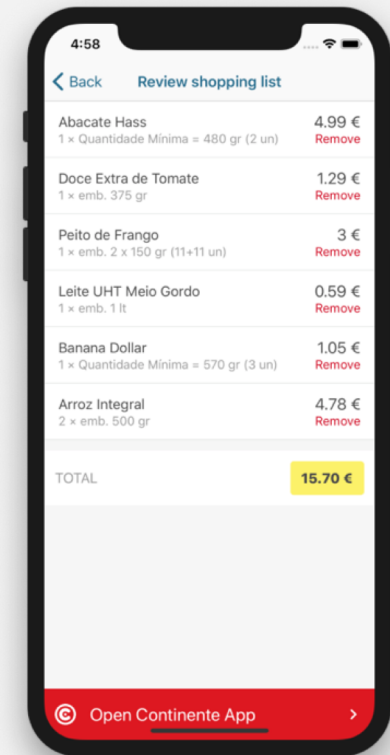
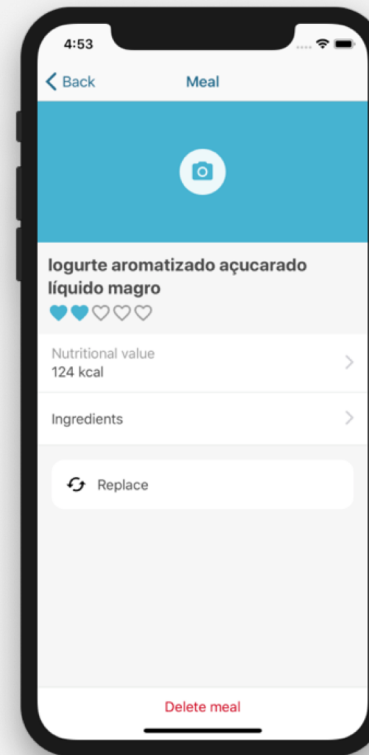
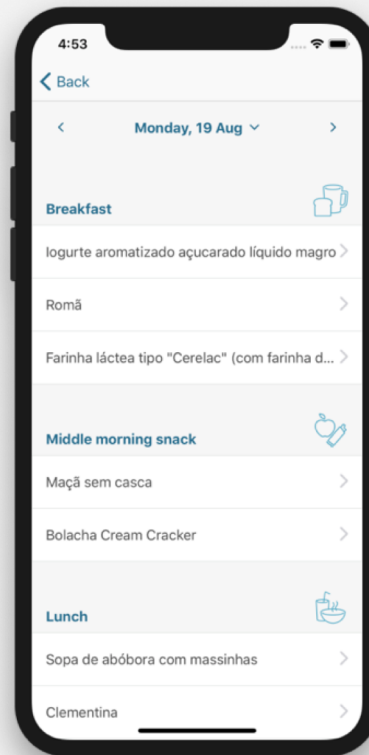
- Education and awareness
 - Change eating habits
- Help users to keep BMI stable at higher age and changing metabolism
 - Prevent malnutrition (over- and undernutrition)

Meal planning instead of food logging

- Many food logging and meal diary apps are available, but
 - It takes time to search databases
 - No precise nutritional values known
- We follow a complementary approach
 - The meal recommender selects recipes
 - We calculate precisely the nutritional values
 - The user just gives simple feedback (meal eaten or not)

Mobile application

- Meal plan
 - Day or week
 - Energy goal
- Recipe details
 - Ingredients
 - Instructions
- E-Commerce
 - Delivery service




Food Information Resources



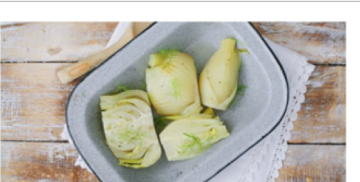
LanguaL thesaurus

| | Serving Size | Calories | Total Fat (g) | Saturated Fat (g) | Cholesterol (mg) | Sodium (mg) | Total Carbohydrate (g) | Dietary Fiber (g) | Sugar (g) | Protein (g) |
|--------------------|--------------|----------|---------------|-------------------|------------------|-------------|------------------------|-------------------|-----------|-------------|
| BALSAMIC | 2 1 oz cont | 195 | 20.74 | 2.85 | 0 | 228.54 | 2.18 | 0 | 1.68 | 0 |
| BLUE CHEESE | 2 1 oz cont | 156 | 13.21 | 3.48 | 15.42 | 361 | 8.19 | 0 | 2.42 | 1.93 |
| RANCH | 2 1 oz cont | 157 | 12.89 | 2.05 | 11 | 388 | 10 | 0.2 | 0 | 1 |
| CAESAR | 2 1 oz cont | 346 | 37.79 | 3.42 | 2.52 | 332 | 0.97 | 0 | 0.4 | 1.74 |
| HONEY MUSTARD | 2 1 oz cont | 306 | 31.28 | 4.8 | 13.8 | 295 | 5.6 | 0.14 | 6.2 | 0.22 |
| ITALIAN (LOW CAL) | 2 1 oz cont | 20 | 0 | 0 | 0 | 1000 | 4 | 0 | 2 | 0 |
| LEMON HERB | 2 1 oz cont | 197 | 17.97 | 2.59 | 9.77 | 338 | 10.05 | 0 | 2.67 | 0.4 |
| FAT FREE RASPBERRY | 2 1 oz cont | 90 | 0 | 0 | 0 | 540 | 16 | 0 | 16 | 0 |
| ORIENTAL | 2 1 oz cont | 59 | 0 | 0 | 0 | 1150 | 14.98 | 0 | 2.31 | 0.64 |
| PLAIN BALSAMIC | 2 1 oz cont | 40 | 0 | 0 | 0 | 0 | 0 | 0 | 8 | 0 |
| OLIVE OIL | 2 1 oz cont | 480 | 56 | 4 | 0 | 0 | 0 | 0 | 0 | 0 |
| FAT FREE RANCH | 2 1 oz cont | 50 | 0 | 0 | 0 | 500 | 12 | 0 | 4 | 0 |
| 1000 ISLAND | 2 1 oz cont | 240 | 24 | 5 | 18 | 420 | 12 | 0 | 10 | 0 |
| SALSA | 2 1 oz cont | 22 | 0.48 | 0.24 | 1125 | 332.27 | 0.95 | 0.07 | 1.63 | 0.87 |
| CILANTRO | 2 1 oz cont | 114 | 12.24 | 1.8 | 0.8 | 80.4 | 0.55 | 0.43 | 0.21 | 0.82 |
| SOUR CREAM | 2 1 oz cont | 121 | 11.9 | 7.4 | 25 | 30 | 2.4 | 0 | 0 | 1.8 |


EuroFIR FCDBs NL, PT, CH




Prei met geitenkaas en linzen
Een bijzondere combinatie van smaken



Gestooftde venkel
Past goed bij aardappelen en varkensvleesreepjes



Surinaamse kippenstoof
Verrassende variant op bekende klassieker



Pangasius met amandelchermoula en tomatensaus
Vis met een Noord-Afrikaanse twist

300+ Voedingscentrum recipes
30 Allerhande recipes

Semantic food knowledge model



Table 5: Weight yield factors by cooking of meat based dishes - beef, continues

| Kind of food, dish | Raw product | Cooking | Cook Temp range °C | Yield factor | | n |
|---|-------------|---------|--------------------|-----------------|-------------------|-------------|
| | | | | with water (kg) | edible part (%wt) | |
| Pot roast beef ¹⁾ (chuck, leg, shoulder), with gravy - meat part, well done - gravy part | FK m,Z | e/s | - 90 | - | - | 1.08 0.05 3 |
| - | - | - | - | - | - | 0.48 0.05 3 |
| - | - | - | - | - | - | 0.60 0.04 3 |
| Roast beef ²⁾ (chop, rump), medium | FK m,S | d | 66-75 | 0.78 | 0.02 | 0.65 0.02 3 |
| Roast beef ³⁾ (chop, rump), well done | FK m,S | d | - 90 | 0.73 | 0.02 | 0.60 0.02 3 |
| Roast beef ⁴⁾ (sirloin), rare | FK m,F | e/g | 55-65 | - | - | 0.83 0.03 3 |
| Roast beef ⁵⁾ (sirloin), medium | FK m,F | e/g | 66-75 | - | - | 0.75 0.04 6 |
| Roast beef ⁶⁾ (sirloin), well done | FK m,F | e/g | 76-85 | - | - | 0.69 0.03 2 |
| Pot roast beef ⁷⁾ (sirloin), with gravy, medium - meat part - gravy part | FK m,Z | d/s | 66-75 | - | - | 0.99 - 1 |
| - | - | - | - | - | - | 0.45 - 1 |
| Pot roast beef ⁸⁾ (sirloin), with gravy, well done - meat part - gravy part | FK m,Z | d/s | 76-85 | - | - | 0.87 - 1 |
| - | - | - | - | - | - | 0.43 - 1 |
| Steak ⁹⁾ (beef, chuck, shoulder, sirloin), rare | FK m,F | d | 55-65 | - | - | 0.82 - 1 |
| Steak ¹⁰⁾ (beef, chuck, shoulder, sirloin), medium | FK m,F | d | 66-75 | - | - | 0.73 - 1 |
| Steak ¹¹⁾ (beef, chuck, shoulder, sirloin), well done | FK m,F | d | 76-85 | - | - | 0.63 - 1 |
| Beef ¹²⁾ (chuck, shoulder), ground - meat part - sauce part | FK m,Z | s | - 90 | - | - | 0.73 0.02 3 |
| - | - | - | - | - | - | 0.35 0.02 3 |
| - | - | - | - | - | - | 0.38 0.02 3 |
| Beef olive ¹³⁾ (top round), with gravy - meat part - gravy part | FK m,Z | s | - 90 | - | - | 0.94 0.02 3 |
| - | - | - | - | - | - | 0.47 0.02 3 |
| - | - | - | - | - | - | 0.47 0.02 3 |
| Spiced vinegar marinated beef ¹⁴⁾ (chuck), with gravy - meat part - gravy part | FK m,Z | s | - 90 | - | - | 0.79 0.02 3 |
| - | - | - | - | - | - | 0.29 0.02 3 |
| - | - | - | - | - | - | 0.50 0.02 3 |

X = mean value of available data, s = confidence interval (p = 0.05); - = not available or not known
n = number of available data, FK = fork, rare¹⁾ = big piece (> 5 kg, high > 2.5 cm), ²⁾ = small piece (20-500 g, high < 2.5 cm),
a = boil, pressure boil, b = steam, pressure steam, c = stew, pressure stew, microwave stew, d = fry in pan with fat, e = fry in oven (dry oil), f = deep fry, g = grill, broil, barbecue, s = braise, m,Z = with ingredients - common German recipe,
a,Z = without any ingredients, m,S = with salt and spice, m,F = with ingredients and fat absorption, common German recipe

Yield and retention factors

Food ontologies

- Comparison between food ontologies (SMAP'2018)

Table 1: Comparison between food ontologies.

| Name | Person | Disease | Recipe | Nutrition | Multilingual | Availability |
|--------------|--------|---------|--------|-----------|--------------|--------------|
| Food Product | | ~ | | X | | X |
| LIRMM | | | | X | | X |
| FOODS | | X | X | X | | |
| Edamam | | X | X | X | | |
| FOKB | X | X | | | | |
| FoodOn | | | X | X | | X |
| FOD | | X | X | X | | |
| OFS | | | X | | | |
| AGROVOC | | | | | X | X |
| LanguaL | | | | | X | X |



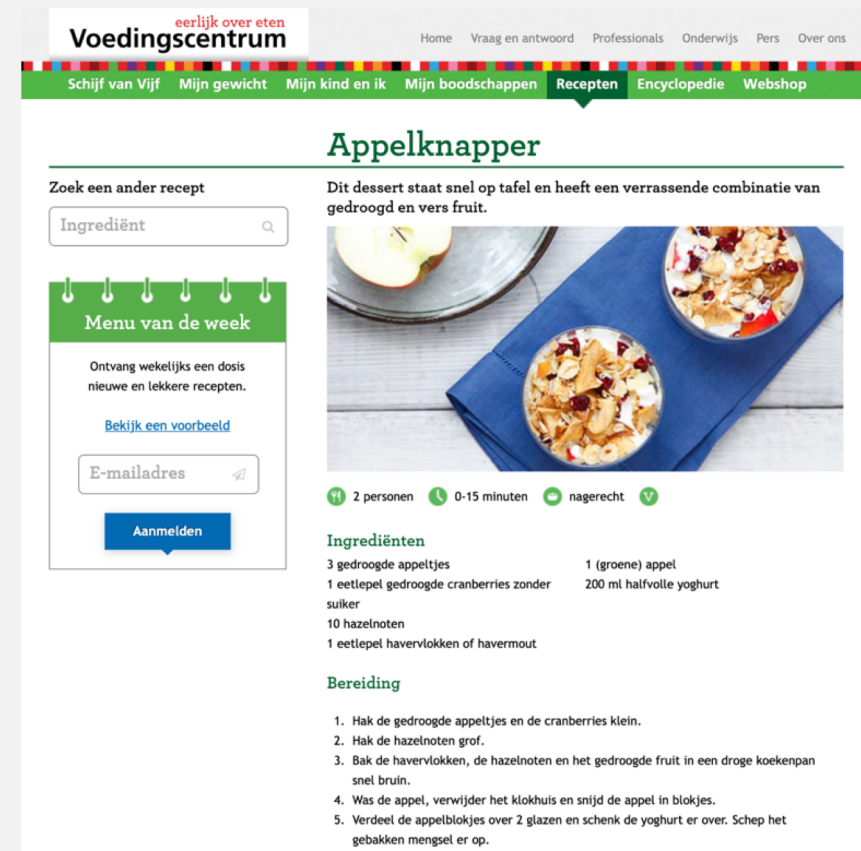
Andreas Arens-Volland, Benjamin Gateau and Yannick Naudet. **Semantic Modeling for Personalized Dietary Recommendation**. In: Proceedings of the 13th International Workshop on Semantic and Social Media Adaptation and Personalization (SMAP2018), Zaragoza, Spain.

- Different facets describe different aspects of a food:
 - **Facet A: PRODUCT TYPE**
 - **Facet B: FOOD SOURCE**
 - **Facet C: PART OF PLANT OR ANIMAL**
 - Facet E: PHYSICAL STATE, SHAPE OR FORM
 - Facet F: EXTENT OF HEAT TREATMENT
 - Facet G: COOKING METHOD
 - Facet H: TREATMENT APPLIED
 - Facet J: PRESERVATION METHOD
 - Facet K: PACKING MEDIUM
 - Facet M: CONTAINER OR WRAPPING
 - Facet N: FOOD CONTACT SURFACE
 - Facet P: CONSUMER GROUP/DIETARY USE/LABEL CLAIM
 - Facet R: GEOGRAPHIC PLACES AND REGIONS
 - Facet Z: ADJUNCT CHARACTERISTICS OF FOOD

languaL.org

Use case 1: Integration of external recipe sources

1. Parse list of ingredients
 1. Quantity
 2. Unit
 3. Food name
2. Find matching foods in FCDB
3. Manual validation of food references
4. Convert kitchen units to grams



The screenshot shows the Voedingscentrum website interface. At the top, there is a navigation bar with links for Home, Vraag en antwoord, Professionals, Onderwijs, Pers, and Over ons. Below this is a secondary navigation bar with links for Schijf van Vijf, Mijn gewicht, Mijn kind en ik, Mijn boodschappen, Recepten, Encyclopedie, and Webshop. The main content area features a recipe for 'Appelknapper'. The recipe title is 'Appelknapper' in green. Below the title, there is a search bar for 'Zoek een ander recept' with a search box labeled 'Ingrediënt'. To the right of the search bar, there is a description: 'Dit dessert staat snel op tafel en heeft een verrassende combinatie van gedroogd en vers fruit.' Below the description is a photograph of the recipe. Under the photo, there are icons for '2 personen', '0-15 minuten', 'nagerecht', and a green checkmark. Below the photo, there is a section for 'Ingrediënten' with a list of ingredients: 3 gedroogde appeltjes, 1 eetlepel gedroogde cranberries zonder suiker, 10 hazelnoten, 1 eetlepel havervlokken of havermout, and 1 (groene) appel, 200 ml halfvolle yoghurt. Below the ingredients, there is a section for 'Bereiding' with a list of five steps: 1. Hak de gedroogde appeltjes en de cranberries klein. 2. Hak de hazelnoten grof. 3. Bak de havervlokken, de hazelnoten en het gedroogde fruit in een droge koekenpan snel bruin. 4. Was de appel, verwijder het klokhuis en snijd de appel in blokjes. 5. Verdeel de appelblokjes over 2 glazen en schenk de yoghurt er over. Schep het gebakken mengsel er op.

Voedingscentrum.nl

Matching of ingredients with foods in EuroFIR



EuroFIR uses special syntax for food names

- Paprika groene rauw = Groene paprika
- Selderij knol- gekookt = Knolselderij
- Vis vet >5 g vet gem rauw

Problems related to text distance matching

- Synonyms
- Additional properties, keywords
- Abbreviations

Involve human experts in validation



| recipe | url | ingredient | NEVO | comment | m1 | m2 | m3 | m4 | m5 |
|---|---|---|---------------------|---------|----------------------|----------------------------|----------------------------------|-------------------------------|------|
| Tiramisu met fruit | https://www.voedingscentrum.nl/recepten/gezond-recept/tiramisumetfruit.aspx | | | | | | | | |
| | | 100 gram bramen | | | Bramen | | | | |
| | | 2 lange vingers | | | Lange vingers | | | | |
| | | 100 ml magere vanillevla met zoetstof | | | | | | | |
| | | 100 gram aardbeien | | | Aardbeien | Puree aardbeien- blik/glas | Aardbeien op siroop blik/glas | Aardbeien op siroop blik/glas | |
| | | 1 sinaasappel | Sinaasappel | | Sinaasappel | Koekje PiM's sinaasappel | Sap sinaasappel- gepasteuriseerd | Sap sinaasappel- vers geperst | Sap |
| | | 50 gram blauwe bessen | | | Bessen blauwe | Bessen zwarte | Bessen rode | Bessen bos- | Bes |
| | | 100 ml halfvolle yoghurt | Yoghurt halfvolle | | Yoghurt halfvolle | Yoghurt vanille- halfvolle | Yoghurt halfvolle Vifit naturel | Yoghurt halfvolle lactosevrij | Yog |
| Couscous met gegrilde vis en groente | https://www.voedingscentrum.nl/recepten/gezond-recept/couscousmetgegrildevisengroente.aspx | | | | | | | | |
| | | 1 courgette | Courgette rauw | | Courgette gekookt | Courgette rauw | | | |
| | | 1 aubergine | | | Aubergine gekookt | Aubergine gekookt | Aubergine rauw | Aubergine rauw | Aut |
| | | 1 gele paprika | | | Paprika gele gekookt | Paprika gele rauw | Fritessticks paprika | Paprika groene gekookt | Paq |
| | | 3 eetlepels olie | Olie zonnebloem- | | Olie frituur- | Olie arachide- | Olie soja- | Olie zonnebloem- | Olie |
| | | ¼ bouillontablet met minder zout | | | | | | | |
| | | ½ theelepel komijnzaad | Komijnzaad gedroogd | | Komijnzaad gedroogd | | | | |
| | | 1 ui | Ui rauw | | Ui gekookt | Ui gebakken | Ui rauw | Ui sla- rauw | Ui z |
| | | 150 gram volkoren couscous | | | Couscous gekookt | Couscous rauw | Knackebrod volkoren | Cracottes volkoren | Bro |
| | | 1 teentje knoflook | Knoflook rauw | | Knoflook rauw | Saus knoflook- 20% olie | | | |
| | | 2 stukjes vis, bijv. kabeljauw, koolvis | | | | | | | |
| | | 2 - 3 theelepels ras el hanout | | | | | | | |
| Kruidig gevulde makreel | https://www.voedingscentrum.nl/recepten/gezond-recept/kruidiggevuudemakreel.aspx | | | | | | | | |
| | | 1 eetlepel kokosmelk (blik) | | | Kokosmelk | IJs op basis van kokosmelk | | | |
| | | 1 theelepel citroensap | Sap citroen- vers | | Sap citroen- vers | | | | |
| | | 1 eetlepel olie | Olie zonnebloem- | | Olie frituur- | Olie arachide- | Olie soja- | Olie zonnebloem- | Olie |
| | | 2 kleine makrelen | | | | | | | |

Use case 2: Alignment between national FCDBs

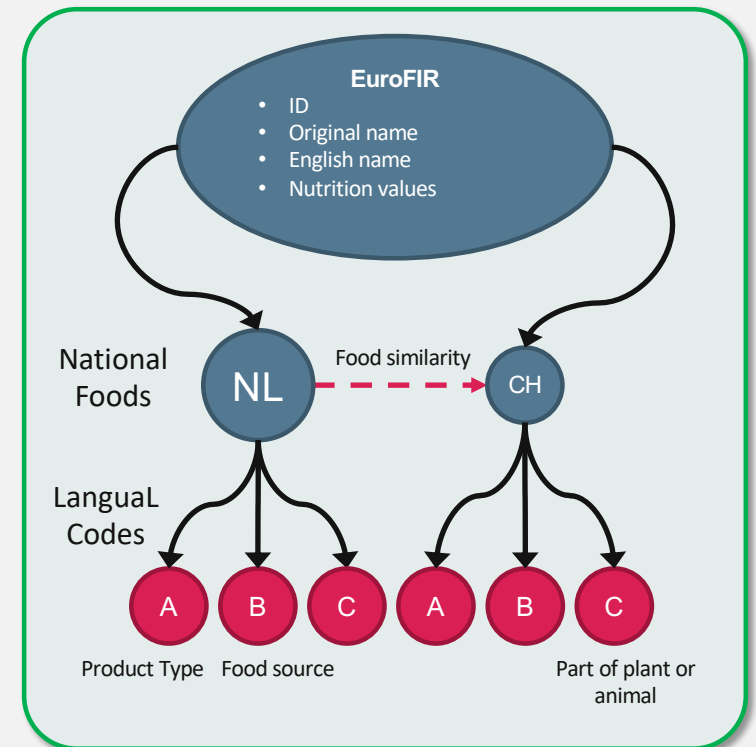
- **Localisation of recipes** requires translation and use of national FCDB

 NEVO Database by RVIM

 Swiss Food Composition Database by FSVO

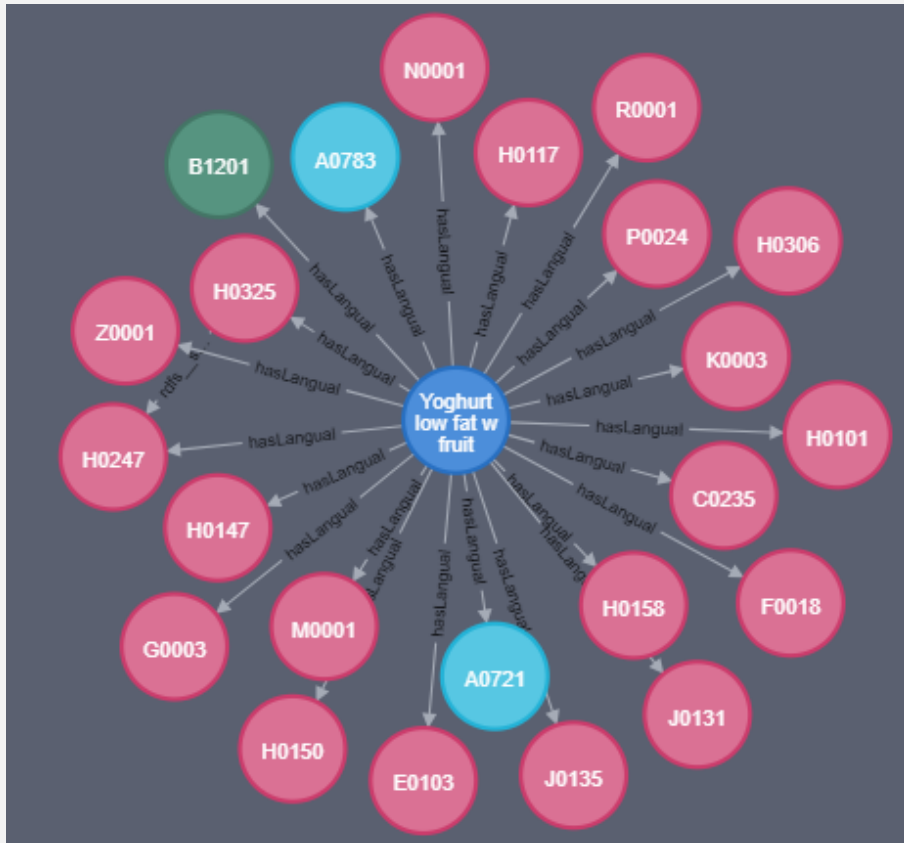
- But: no mappings available!

- Approach:
Semi-automatically create missing links by computing a hybrid syntactic and semantic similarity measure.



Food example

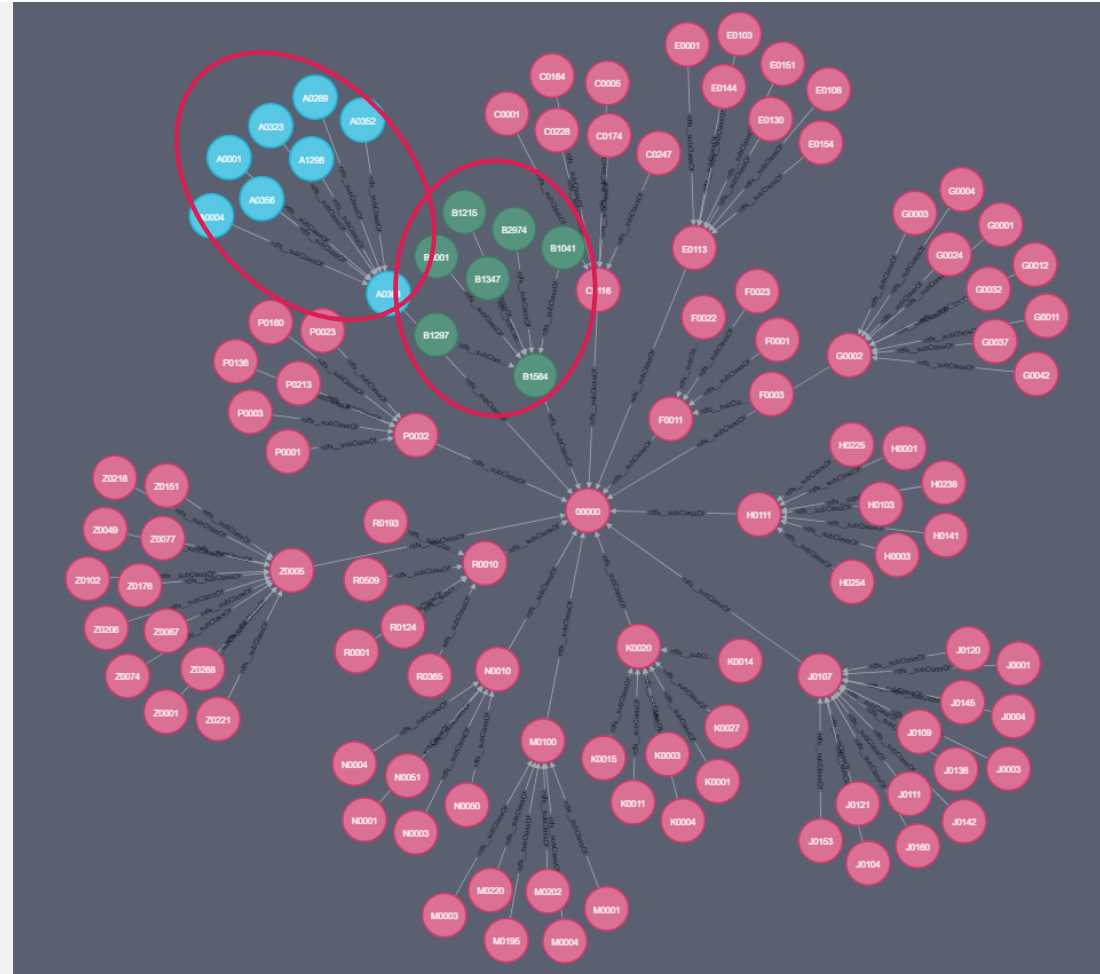
“Yoghurt low fat w fruit”



| LanguaLID | LanguaLabel |
|-----------|--|
| "A0721" | "31 OTHER MILK PRODUCTS (EFG)" |
| "A0783" | "FERMENTED MILK PRODUCT (EUROFIR)" |
| "B1201" | "COW" |
| "C0235" | "MILK" |
| "E0103" | "SEMILIQUID" |
| "F0018" | "PARTIALLY HEAT-TREATED" |
| "G0003" | "COOKING METHOD NOT APPLICABLE" |
| "H0101" | "LACTIC ACID FERMENTED" |
| "H0117" | "FLAVORING OR TASTE INGREDIENT ADDED" |
| "H0147" | "FRUIT ADDED" |
| "H0150" | "COLOR ADDED" |
| "H0158" | "SUCROSE ADDED" |
| "H0247" | "FAT PARTIALLY REMOVED" |
| "H0306" | "HOMOGENIZED OR EMULSIFIED" |
| "H0325" | "FAT PARTIALLY REMOVED, LESS THAN 50% REMAINING" |
| "J0131" | "PRESERVED BY CHILLING" |
| "J0135" | "PASTEURIZED BY HEAT" |
| "K0003" | "NO PACKING MEDIUM USED" |
| "M0001" | "CONTAINER OR WRAPPING NOT KNOWN" |
| "N0001" | "FOOD CONTACT SURFACE NOT KNOWN" |
| "P0024" | "HUMAN CONSUMER, NO AGE SPECIFICATION" |
| "R0001" | "GEOGRAPHIC PLACE OR REGION NOT KNOWN" |
| "Z0001" | "ADJUNCT CHARACTERISTICS OF FOOD NOT KNOWN" |

LanguaL tree structure

- Tree visualisation
 - A PRODUCT TYPE
 - B FOOD SOURCE



Food matching

“Yoghurt low fat w fruit”



VS



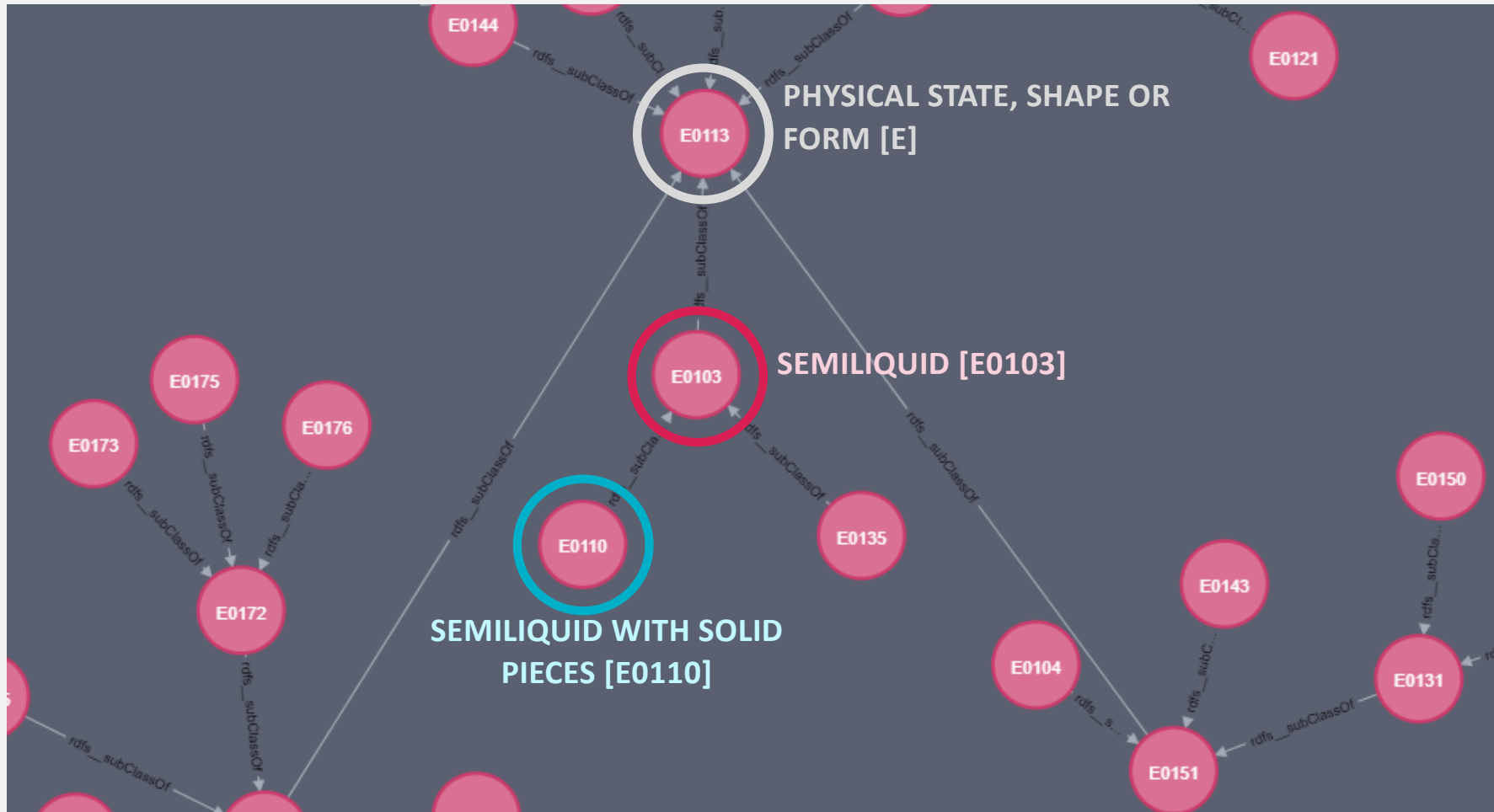
“Yogurt with fruit, low fat, with sweetener”

| LanguaLID | LanguaL label |
|----------------|--|
| "A0721" | "31 OTHER MILK PRODUCTS (EFG)" |
| "A0783" | "FERMENTED MILK PRODUCT (EUROFIR)" |
| "B1201" | "COW" |
| "C0235" | "MILK" |
| "E0103" | "SEMILIQUID" |
| "F0018" | "PARTIALLY HEAT-TREATED" |
| "G0003" | "COOKING METHOD NOT APPLICABLE" |
| "H0101" | "LACTIC ACID FERMENTED" |
| "H0117" | "FLAVORING OR TASTE INGREDIENT ADDED" |
| "H0147" | "FRUIT ADDED" |
| "H0150" | "COLOR ADDED" |
| "H0158" | "SUCROSE ADDED" |
| "H0247" | "FAT PARTIALLY REMOVED" |
| "H0306" | "HOMOGENIZED OR EMULSIFIED" |
| "H0325" | "FAT PARTIALLY REMOVED, LESS THAN 50% REMAINING" |
| "J0131" | "PRESERVED BY CHILLING" |
| "J0135" | "PASTEURIZED BY HEAT" |
| "K0003" | "NO PACKING MEDIUM USED" |
| "M0001" | "CONTAINER OR WRAPPING NOT KNOWN" |
| "N0001" | "FOOD CONTACT SURFACE NOT KNOWN" |
| "P0024" | "HUMAN CONSUMER, NO AGE SPECIFICATION" |
| "R0001" | "GEOGRAPHIC PLACE OR REGION NOT KNOWN" |
| "Z0001" | "ADJUNCT CHARACTERISTICS OF FOOD NOT KNOWN" |



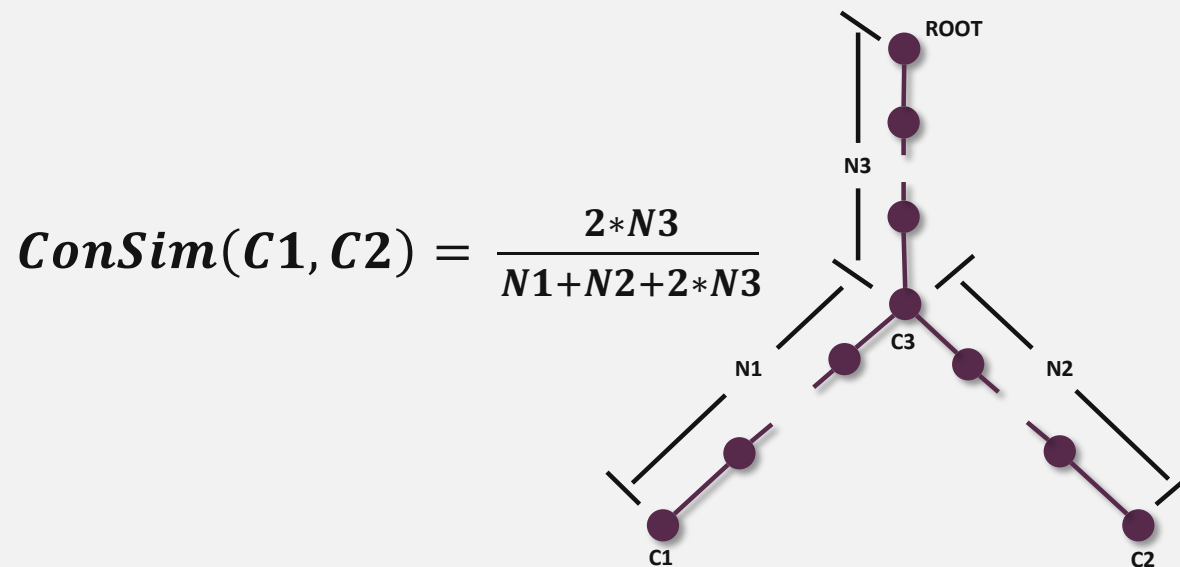
| LanguaLID | LanguaL label |
|----------------|---|
| "A0101" | "CULTURED MILK PRODUCT (US CFR)" |
| "A0783" | "FERMENTED MILK PRODUCT (EUROFIR)" |
| "B1201" | "COW" |
| "C0235" | "MILK" |
| "E0110" | "SEMILIQUID WITH SOLID PIECES" |
| "F0018" | "PARTIALLY HEAT-TREATED" |
| "G0003" | "COOKING METHOD NOT APPLICABLE" |
| "H0101" | "LACTIC ACID FERMENTED" |
| "H0108" | "NONNUTRITIVE SWEETENER ADDED" |
| "H0147" | "FRUIT ADDED" |
| "J0131" | "PRESERVED BY CHILLING" |
| "J0135" | "PASTEURIZED BY HEAT" |
| "K0003" | "NO PACKING MEDIUM USED" |
| "M0184" | "PLASTIC CONTAINER, RIGID OR SEMIRIGID" |
| "N0001" | "FOOD CONTACT SURFACE NOT KNOWN" |
| "P0024" | "HUMAN CONSUMER, NO AGE SPECIFICATION" |
| "P0040" | "REDUCED FAT FOOD" |
| "R0188" | "SWITZERLAND" |
| "Z0112" | "FOOD INDUSTRY PREPARED" |
| "Z0181" | "FAT CONTENT < 1%" |

Conceptual similarity regarding facet E codes



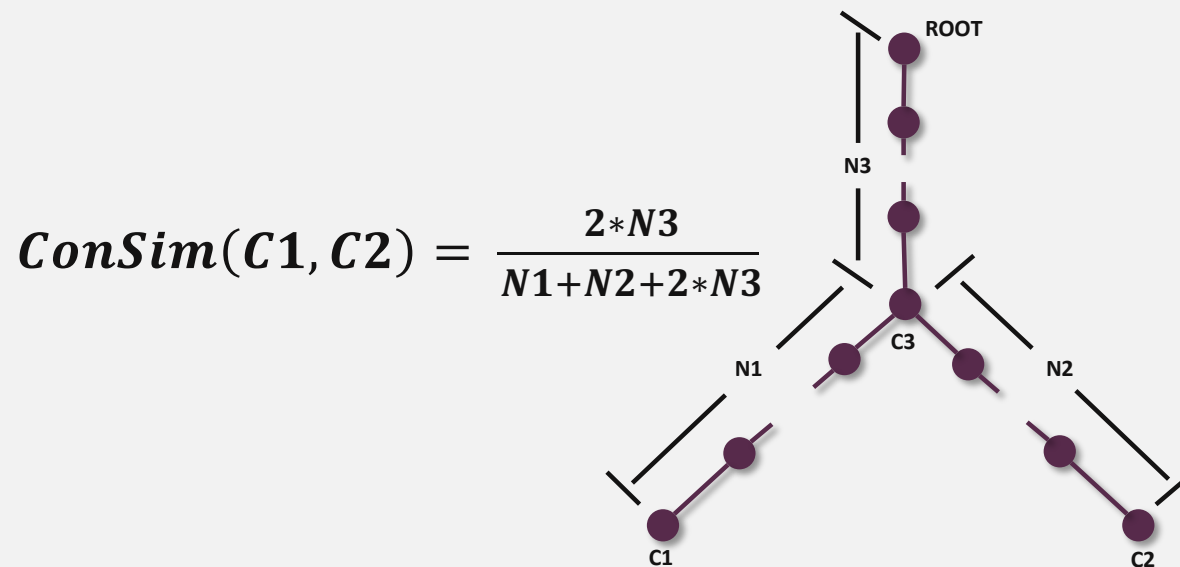
17

Conceptual similarity measure



Z. Wu, M. Palmer (1994). **Verb semantics and lexical selection**, In: 32nd Annual Meeting of the Association for Computational Linguistics, Las Cruces, New Mexico, pp. 133–138, June 1994.

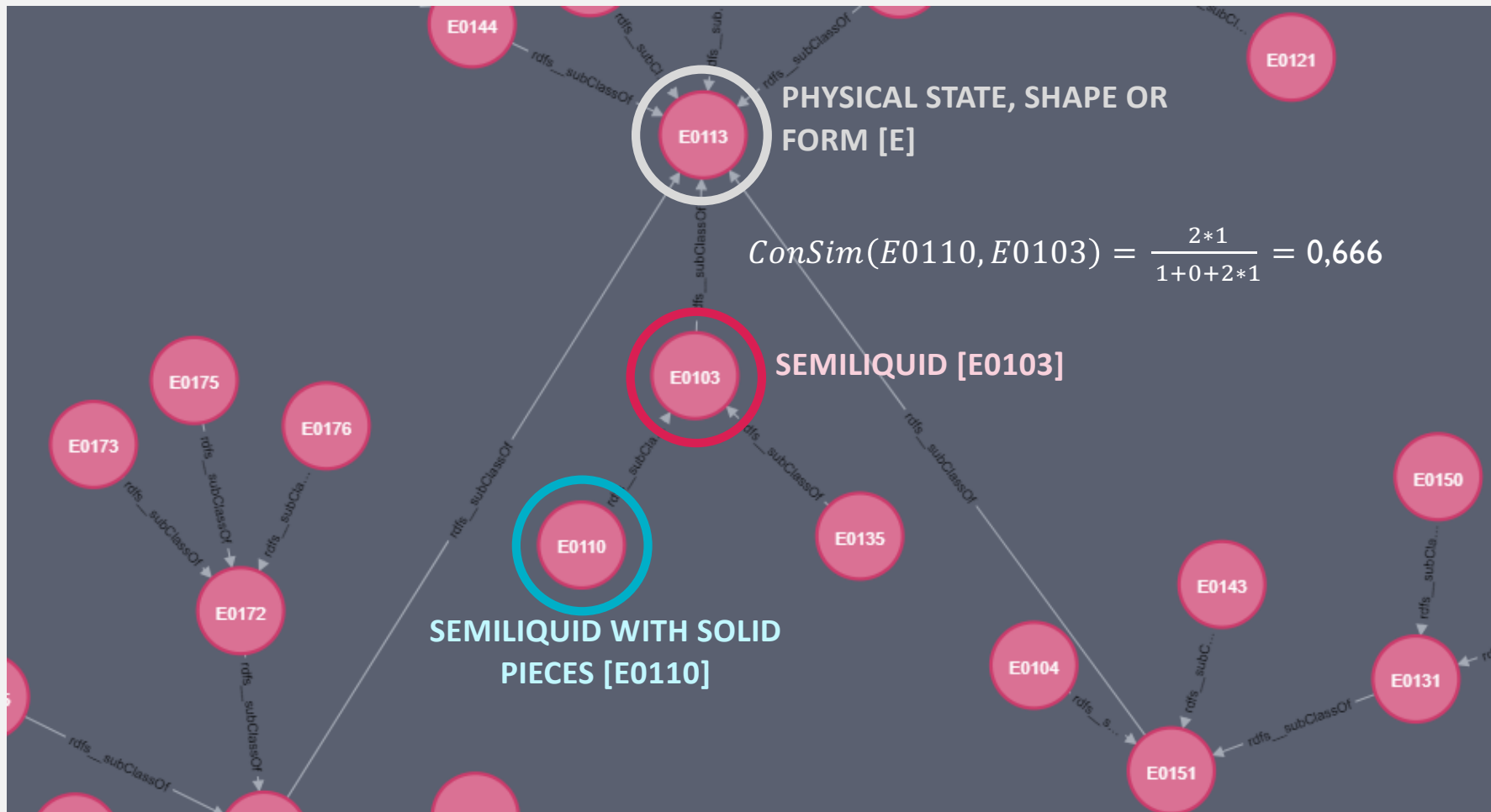
Conceptual similarity measure



In case of multiple measures within a group we maximize. Finally, we average over the similarity of all groups.

Z. Wu, M. Palmer (1994). **Verb semantics and lexical selection**, In: 32nd Annual Meeting of the Association for Computational Linguistics, Las Cruces, New Mexico, pp. 133–138, June 1994.

Conceptual similarity regarding facet E codes



20

Specialisation vs. disjunct concepts

! We can apply the conceptual similarity measure only in parts of the tree where a **subsumption (is-a)** relation is given.

These groups are **disjunct**

- H. TREATMENT APPLIED
 - + COMPONENT REMOVED
 - + COMPONENT SUBSTITUTED
 - FOOD MODIFIED
 - INSTANTIZED
 - MICROBIAALLY/ENZYMATICALLY MODIFIED
 - + ENZYMATICALLY MODIFIED
 - + FERMENTED/MODIFIED, MULTIPLE COMPONENT
 - FERMENTED/MODIFIED, SINGLE COMPONENT
 - CARBOHYDRATE FERMENTED
 - ACETIC ACID FERMENTED
 - ALCOHOL FERMENTED
 - + ALCOHOL-ACETIC ACID FERMENTED
 - LACTIC ACID FERMENTED**
 - LIPOLYTIC FERMENTED
 - PROTEOLYTIC FERMENTED
 - PICKLED
 - + PHYSICALLY/CHEMICALLY MODIFIED
 - RENDERED
 - INGREDIENT ADDED
 - + CARBOHYDRATE OR RELATED COMPOUND ADDED

Same facet H, but disjunct concepts

“Yoghurt low fat w fruit”



VS



“Yogurt with fruit, low fat, with sweetener”

| LanguaLID | LanguaL label |
|----------------|--|
| "A0721" | "31 OTHER MILK PRODUCTS (EFG)" |
| "A0783" | "FERMENTED MILK PRODUCT (EUROFIR)" |
| "B1201" | "COW" |
| "C0235" | "MILK" |
| "E0103" | "SEMILIQUID" |
| "F0018" | "PARTIALLY HEAT-TREATED" |
| "G0003" | "COOKING METHOD NOT APPLICABLE" |
| "H0101" | "LACTIC ACID FERMENTED" |
| "H0117" | "FLAVORING OR TASTE INGREDIENT ADDED" |
| "H0147" | "FRUIT ADDED" |
| "H0150" | "COLOR ADDED" |
| "H0158" | "SUCROSE ADDED" |
| "H0247" | "FAT PARTIALLY REMOVED" |
| "H0306" | "HOMOGENIZED OR EMULSIFIED" |
| "H0325" | "FAT PARTIALLY REMOVED, LESS THAN 50% REMAINING" |
| "J0131" | "PRESERVED BY CHILLING" |
| "J0135" | "PASTEURIZED BY HEAT" |
| "K0003" | "NO PACKING MEDIUM USED" |
| "M0001" | "CONTAINER OR WRAPPING NOT KNOWN" |
| "N0001" | "FOOD CONTACT SURFACE NOT KNOWN" |
| "P0024" | "HUMAN CONSUMER, NO AGE SPECIFICATION" |
| "R0001" | "GEOGRAPHIC PLACE OR REGION NOT KNOWN" |
| "Z0001" | "DISJUNCT CHARACTERISTICS OF FOOD NOT KNOWN" |



| LanguaLID | LanguaL label |
|----------------|---|
| "A0101" | "CULTURED MILK PRODUCT (US CFR)" |
| "A0783" | "FERMENTED MILK PRODUCT (EUROFIR)" |
| "B1201" | "COW" |
| "C0235" | "MILK" |
| "E0110" | "SEMILIQUID WITH SOLID PIECES" |
| "F0018" | "PARTIALLY HEAT-TREATED" |
| "G0003" | "COOKING METHOD NOT APPLICABLE" |
| "H0101" | "LACTIC ACID FERMENTED" |
| "H0108" | "NONNUTRITIVE SWEETENER ADDED" |
| "H0147" | "FRUIT ADDED" |
| "J0131" | "PRESERVED BY CHILLING" |
| "J0135" | "PASTEURIZED BY HEAT" |
| "K0003" | "NO PACKING MEDIUM USED" |
| "M0184" | "PLASTIC CONTAINER, RIGID OR SEMIRIGID" |
| "N0001" | "FOOD CONTACT SURFACE NOT KNOWN" |
| "P0024" | "HUMAN CONSUMER, NO AGE SPECIFICATION" |
| "P0040" | "REDUCED FAT FOOD" |
| "R0188" | "SWITZERLAND" |
| "Z0112" | "FOOD INDUSTRY PREPARED" |
| "Z0181" | "FAT CONTENT < 1%" |

Different facets can be used to model low fat

“Yoghurt low fat w fruit”



VS



“Yogurt with fruit, low fat, with sweetener”

| LanguaLID | LanguaL label |
|-----------|--|
| "A0721" | "31 OTHER MILK PRODUCTS (EFG)" |
| "A0783" | "FERMENTED MILK PRODUCT (EUROFIR)" |
| "B1201" | "COW" |
| "C0235" | "MILK" |
| "E0103" | "SEMILIQUID" |
| "F0018" | "PARTIALLY HEAT-TREATED" |
| "G0003" | "COOKING METHOD NOT APPLICABLE" |
| "H0101" | "LACTIC ACID FERMENTED" |
| "H0117" | "FLAVORING OR TASTE INGREDIENT ADDED" |
| "H0147" | "FRUIT ADDED" |
| "H0150" | "COLOR ADDED" |
| "H0158" | "SUCROSE ADDED" |
| "H0247" | "FAT PARTIALLY REMOVED" |
| "H0306" | "HOMOGENIZED OR EMULSIFIED" |
| "H0325" | "FAT PARTIALLY REMOVED, LESS THAN 50% REMAINING" |
| "J0131" | "PRESERVED BY CHILLING" |
| "J0135" | "PASTEURIZED BY HEAT" |
| "K0003" | "NO PACKING MEDIUM USED" |
| "M0001" | "CONTAINER OR WRAPPING NOT KNOWN" |
| "N0001" | "FOOD CONTACT SURFACE NOT KNOWN" |
| "P0024" | "HUMAN CONSUMER, NO AGE SPECIFICATION" |
| "R0001" | "GEOGRAPHIC PLACE OR REGION NOT KNOWN" |
| "Z0001" | "ADJUNCT CHARACTERISTICS OF FOOD NOT KNOWN" |



| LanguaLID | LanguaL label |
|-----------|---|
| "A0101" | "CULTURED MILK PRODUCT (US CFR)" |
| "A0783" | "FERMENTED MILK PRODUCT (EUROFIR)" |
| "B1201" | "COW" |
| "C0235" | "MILK" |
| "E0110" | "SEMILIQUID WITH SOLID PIECES" |
| "F0018" | "PARTIALLY HEAT-TREATED" |
| "G0003" | "COOKING METHOD NOT APPLICABLE" |
| "H0101" | "LACTIC ACID FERMENTED" |
| "H0108" | "NONNUTRITIVE SWEETENER ADDED" |
| "H0147" | "FRUIT ADDED" |
| "J0131" | "PRESERVED BY CHILLING" |
| "J0135" | "PASTEURIZED BY HEAT" |
| "K0003" | "NO PACKING MEDIUM USED" |
| "M0184" | "PLASTIC CONTAINER, RIGID OR SEMIRIGID" |
| "N0001" | "FOOD CONTACT SURFACE NOT KNOWN" |
| "P0024" | "HUMAN CONSUMER, NO AGE SPECIFICATION" |
| "P0040" | "REDUCED FAT FOOD" |
| "R0188" | "SWITZERLAND" |
| "Z0112" | "FOOD INDUSTRY PREPARED" |
| "Z0181" | "FAT CONTENT < 1%" |

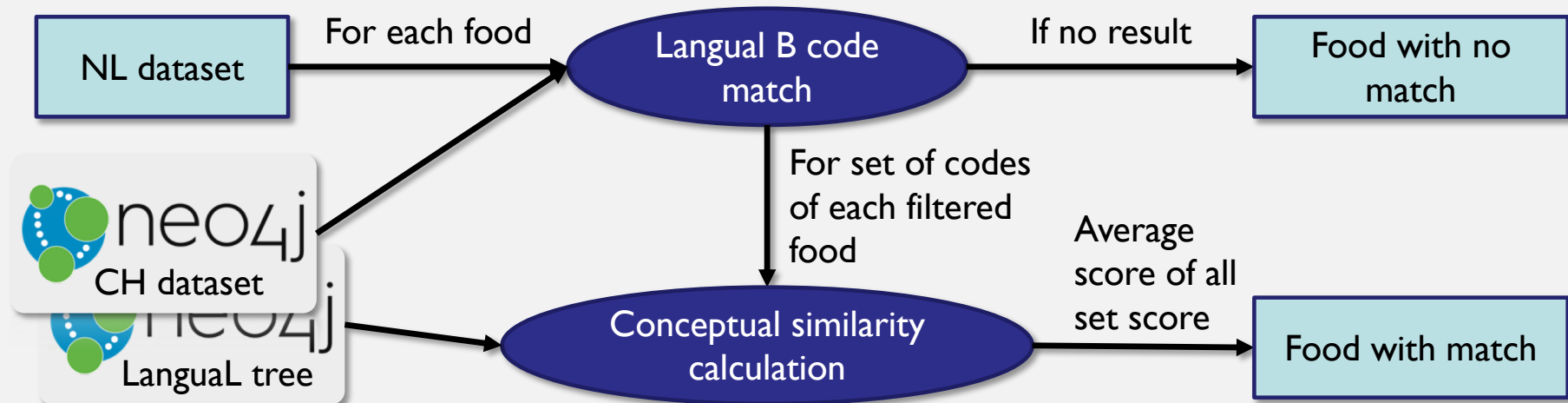
Similarity measuring methods

#1 – **Average** of set of Conceptual Similarities
(for each sub-group, comparison of available LanguaL codes from NL foods and CH foods)

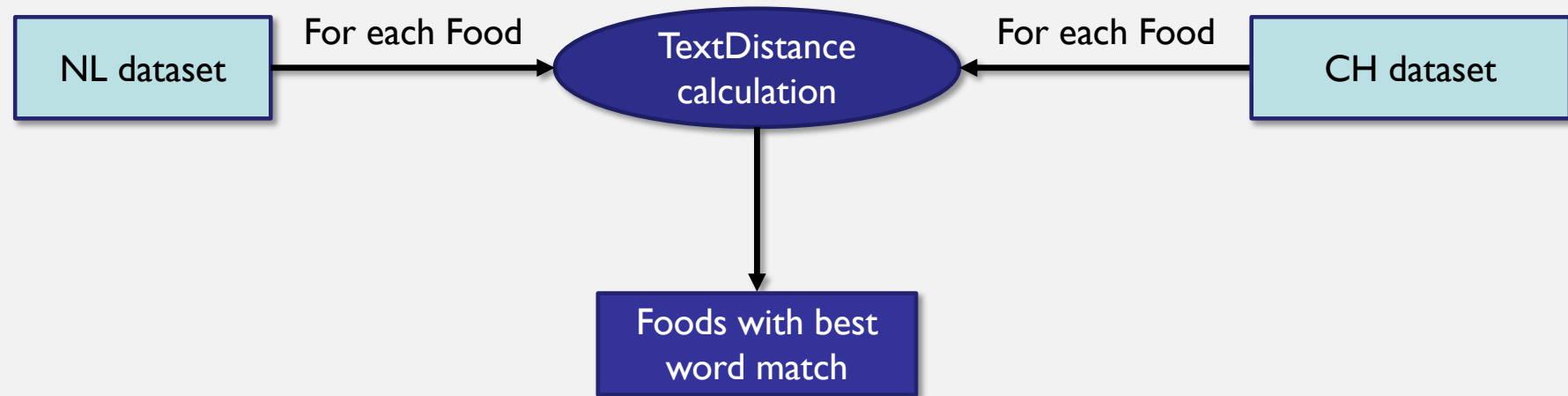
#2 – **Text distance**
(calculation of text distance between English food names in NL and CH datasets)

#3 – **Hybrid version**
(combine text distance and conceptual similarities)

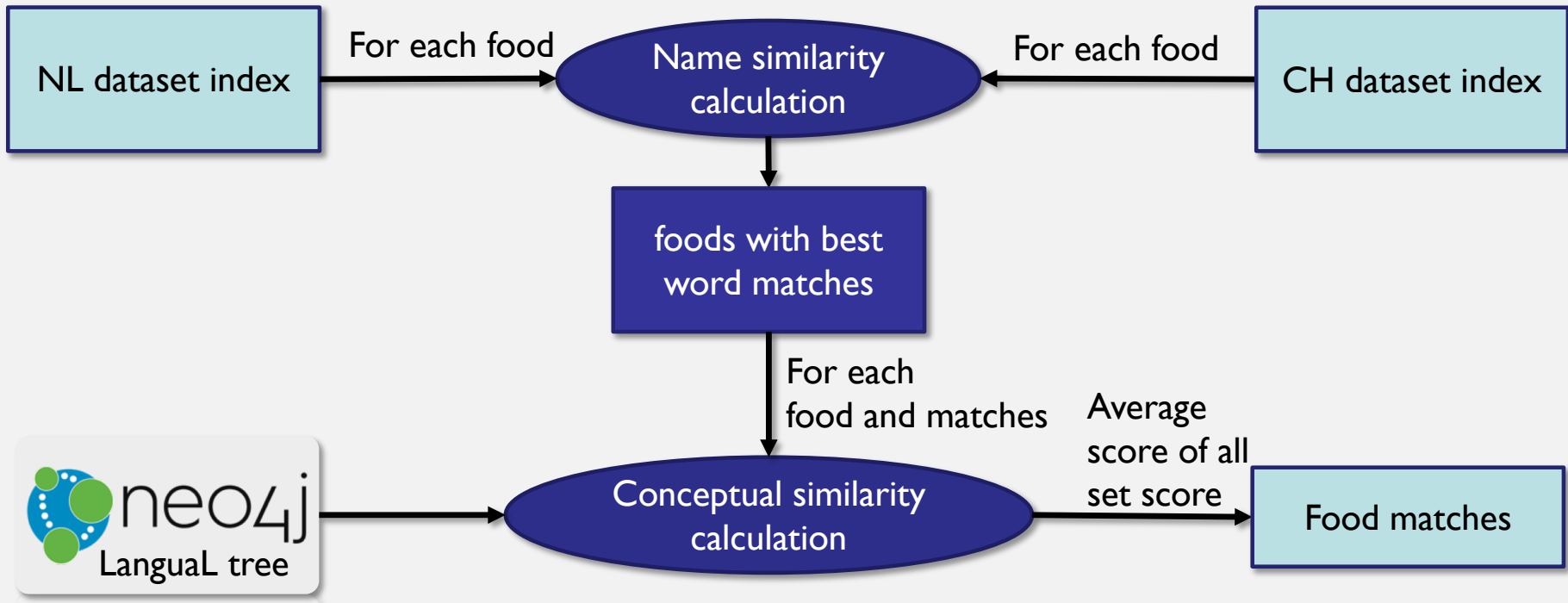
Method #1: Conceptual similarity



Method #2: Text distance



Method #3: Hybrid solution



Examples

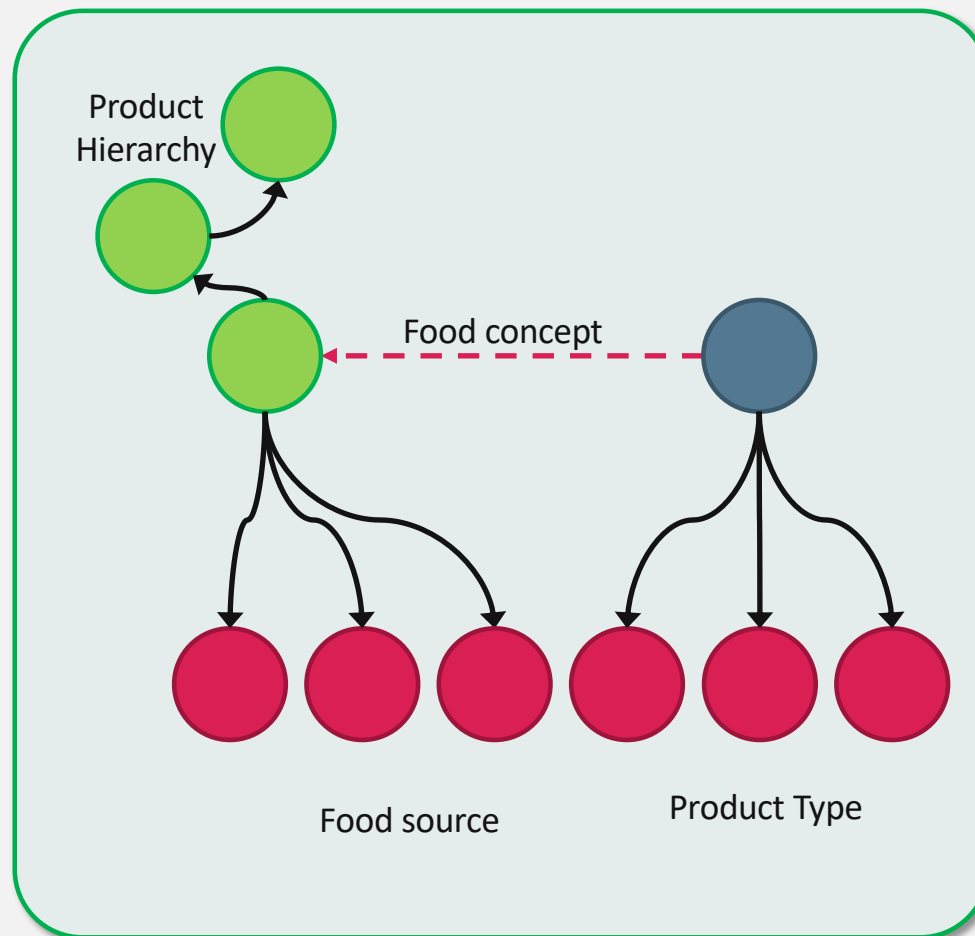
| EName1 | Text distance | Similarity Measure | Average | EName2 |
|--------------------------|---------------|--------------------|--------------|---|
| Yoghurt half fat w fruit | 0,435 | 0,900 | 0,668 | Yogurt |
| Yoghurt half fat w fruit | 0,414 | 0,909 | 0,662 | Yogurt, vanilla |
| Yoghurt half fat w fruit | 0,414 | 0,909 | 0,662 | Yogurt, chocolate |
| Yoghurt half fat w fruit | 0,395 | 0,900 | 0,648 | Yogurt, low fat |
| Yoghurt half fat w fruit | 0,414 | 0,855 | 0,634 | Yogurt (Migros) |
| Yoghurt half fat w fruit | 0,235 | 0,917 | 0,576 | Yogurt, strawberry |
| Yoghurt half fat w fruit | 0,235 | 0,909 | 0,572 | Yogurt, mocca |
| Yoghurt half fat w fruit | 0,235 | 0,909 | 0,572 | Yogurt, hazelnut |
| Yoghurt half fat w fruit | 0,229 | 0,909 | 0,569 | Yogurt, mocca, organic |
| Yoghurt half fat w fruit | 0,212 | 0,917 | 0,564 | Yogurt with fruit, low fat, with sweetener |
| Yoghurt half fat w fruit | 0,153 | 0,889 | 0,521 | Yogurt flavored, low fat, with sweetener |

Preliminary results for category of milk products



| | LanguaL | Hybrid |
|---|----------------|---------------|
| Number of foods | 201 | 201 |
| Number of foods with no match: | 16 | 57 |
| Number of foods with one or several matches: | 185 | 144 |
| Correct matches | 26 | 25 |
| Correct food but wrong preparation method | 83 | 88 |
| Incorrect matches | 76 | 24 |

Future work: Align EuroFIR with FoodOn ontology



Conclusions

- Alignment between food information resources works best with a hybrid approach that combines text distance with conceptual similarity, but results are below our expectations.
- Advantages of the LanguaL thesaurus:
 - Ability to classify foods by various characteristics, such as their source, packaging, and treatment.
 - Supported by EuroFIR food composition datasets and some formal ontologies.
- Disadvantages:
 - Missing guidelines to choose among codes along the hierarchy.
 - Undefined relationship between concepts (subsumption or disjunction).
- Future work
 - Improve text distance by recognition of synonyms (eggplant/aubergine) and typical expressions (“low fat”).
 - Full annotation of disjunct group nodes (requires expert domain knowledge).

Thank you.



Project LIFANA

LIFANA.EU

christoph.stahl@list.lu

Co-funded by the AAL Programme, EU
and the Luxembourg National Research Fund



Luxembourg National
Research Fund