EFSA FoodEx2:
Food Classification System &
Smart Coding App (SCA)

Anastasia LIVANIOU
Scientific Officer
Data Gateway & Outreach team
iDATA Unit

EUROPEAN FOOD SAFETY AUTHORITY

One of the EU bodies that are **responsible for food safety**

The **EU reference body for risk assessment** of food and feed. Its work covers the entire food chain – from field to fork

DATA COLLECTION & STANDARDISATION

- **Data Collections**

  - Biological Monitoring - zoonoses, avian infl., AMR, TSEs, ASF *etc.*

  - Chemical Monitoring - pesticides, additives, contaminants, vet. drugs *etc.*

  - Food Consumption

  - Food Composition

- **Data Standards**

  - Reporting guidance documents

  - Data models (e.g. SSD2)

  - **Controlled Terminology** (e.g. FoodEx2)

  - Data Collection Framework

  - Data validation

- **Data Sources**

  - Member States' monitoring or surveillance programmes

  - Stakeholders (research inst., academia, industry *etc.*)

WHY CONTROLLED TERMINOLOGY *for foods*

Different domain/Regulations

data on chemicals, biological agents, food consumption etc.
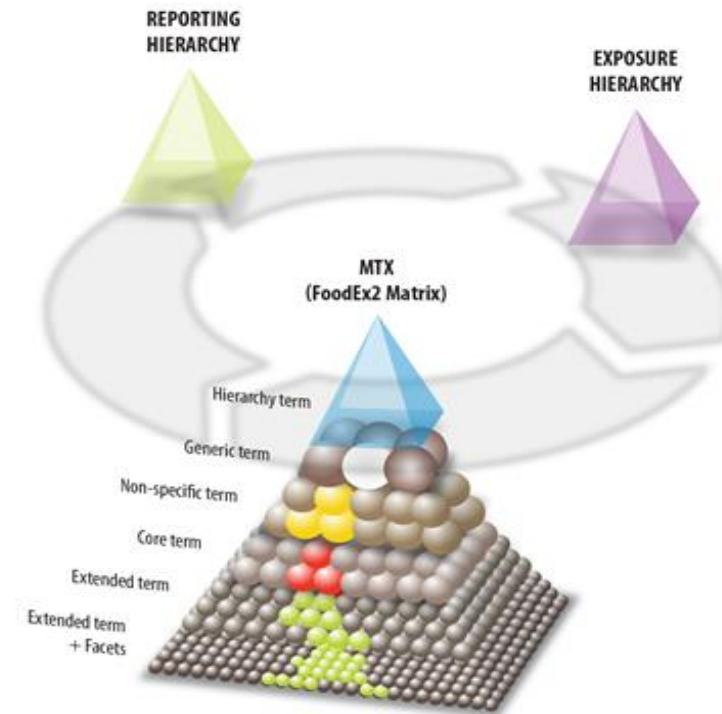
Different samples

data for food, feed, or environmental samples (live animals, soil, plant, etc.)

Different needs

samples described to finest detail or foods aggregated to broader groups

A way to connect different perspectives

- A hierarchical **food/feed classification** and description system

- Harmonisation using a **common language**

USE OF FOODEX2

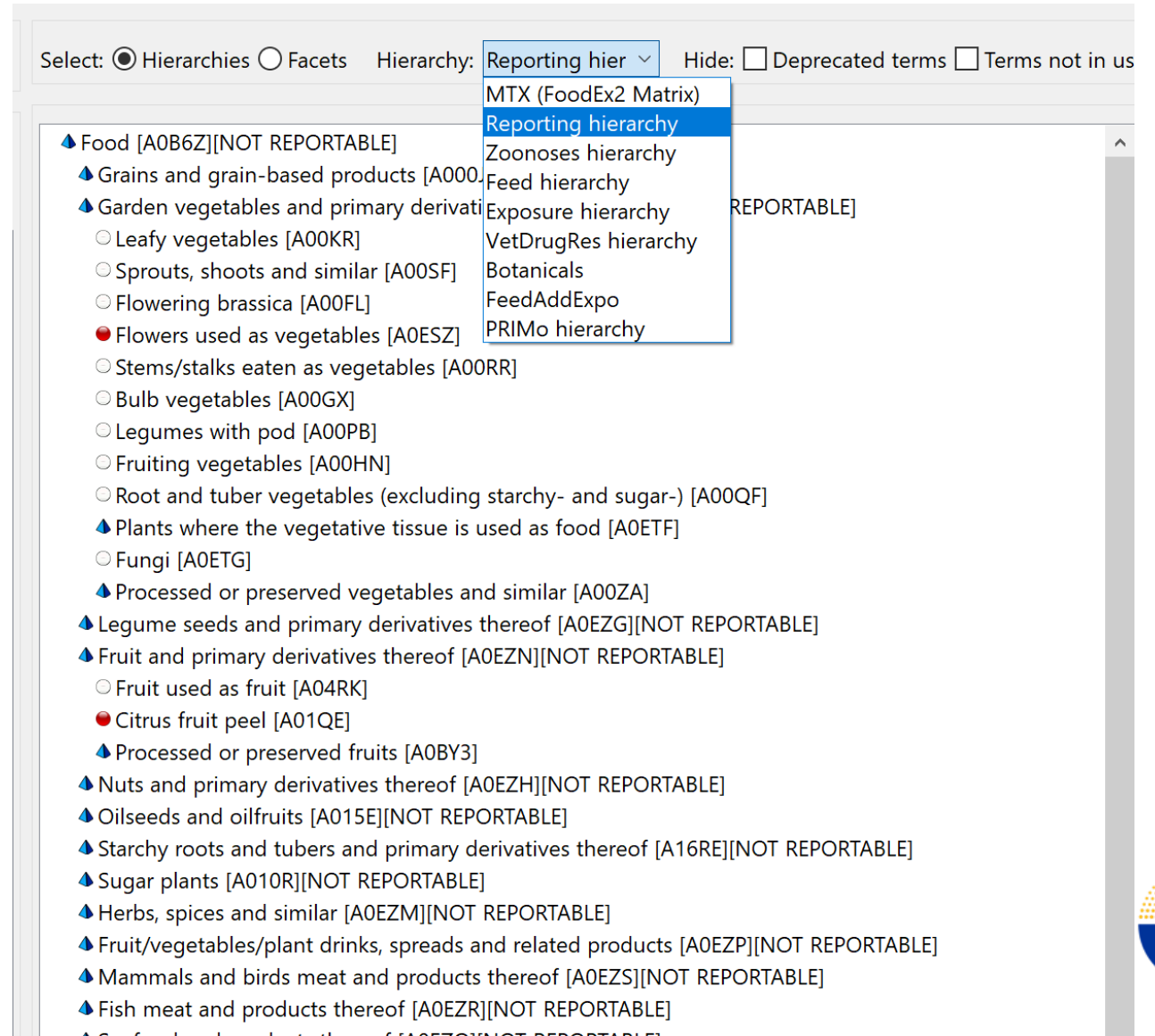Developed and maintained by **EFSA** (3-4 releases/year)

- Used by **EU Member States** - EFSA data collections

- **FAO, WHO** - GIFT database

- **Tufts University** - Global Dietary Database (GDD).

CATALOGUE BROWSER

- **Java application** connected with EFSA master repository

- Supports **all EFSA catalogues**

- Entries aggregated in **parent-child (tree) structure**
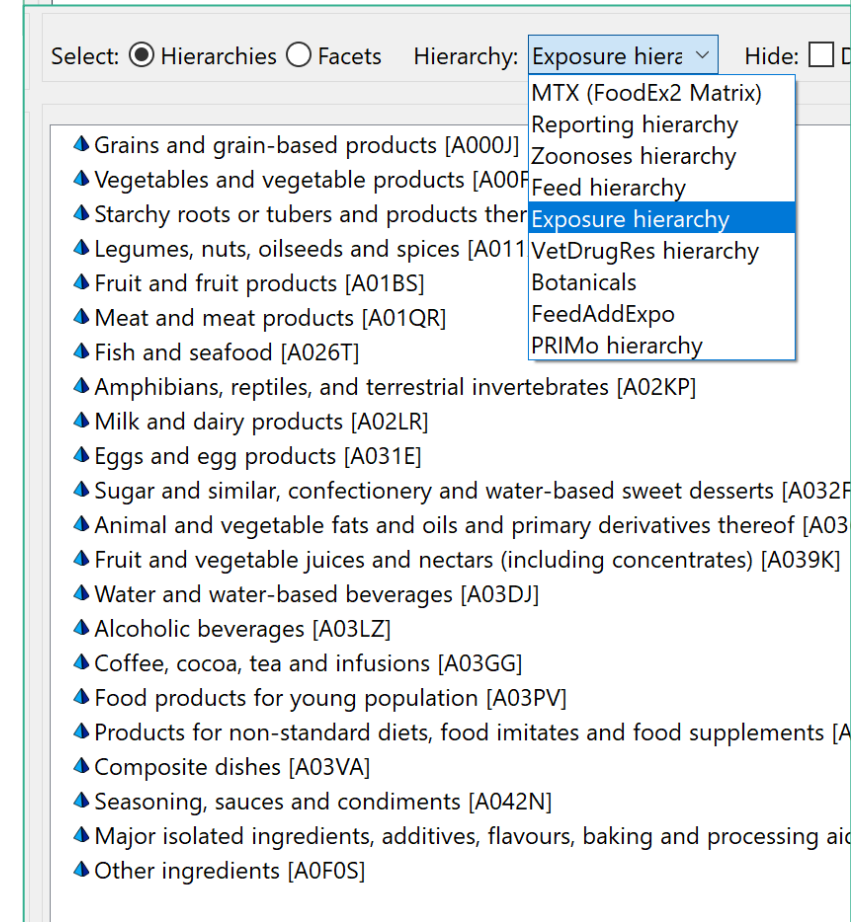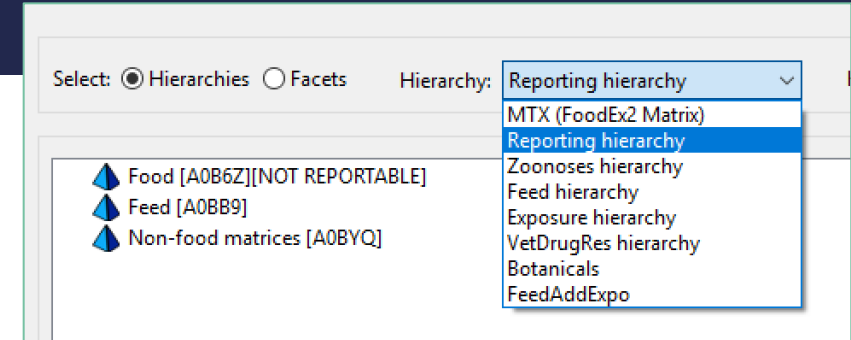
- **Freely available** – OpenAPI

https://github.com/openefsa/catalogue-browser/wiki

FOODEX2 HIERARCHIES

Catalogue host **several hierarchies** used for different data collection

- MTX master hierarchy (ALL terms)

- Reporting hierarchy (chemical occurrence data)

27 food groups in total

- Exposure hierarchy (consumption data)

22 food groups in total

BASE TERMS

Type of term
- Raw Primary Commodity
- RPC derivative
- Composite food simple
- Composite food aggregated (recipes)

Level of detail

▲	Hierarchy term
○	Generic term
● (yellow)	Non-specific term
● (red)	Core term
● (green)	Extended term

BASE TERMS

Grains and grain-based products [A000J]
Vegetables and vegetable products [A00FJ]
Starchy roots or tubers and products thereof, sugar plants [A00ZR]
Legumes, nuts, oilseeds and spices [A011X]
Fruit and fruit products [A01BS]
Meat and meat products [A01QR]
Fish and seafood [A026T]
  Fish (meat) [A026V]
    Freshwater fish [A026X]
    Diadromous fish [A028E]
    Marine fish [A029R]
      Flounders, halibuts, soles [A02AR]
        Dab or common dab [A02AS]
        Flounders [A02AT]
        Halibut [A02AY]
          Halibut, atlantic [A02AZ]
          Halibut, greenland [A02BA]
          Halibut, pacific [A02BB]
        Plaice [A02BC]
        Sole [A02BF]
        Yellowfin sole [A0FAK]
        Turbot [A02BG]
        Brill [A02BH]
        Megrims [A0FAL]
      Cods, hakes, haddocks [A02BT]
      Miscellaneous coastal marine fishes [A029S]
      Miscellaneous demersal marine fishes [A02AG]
      Herrings, sardines, anchovies [A0F9V]
      Tunas, bonitos, billfishes [A0F9X]
      Miscellaneous pelagic marine fishes [A02BM]
      Sharks, rays, chimaeras [A0F9Y]
      Marine fishes not identified [A04RS][DISMISSED]
  Fish offal [A02EH]
  Crustaceans [A02FD]
  Molluscs [A02GM]
  Seafood offal [A16FR]
Sea-squirts and other tunicates [A02GN]

Term naming and definition | Implicit facets | Reportability

Type of term
Raw Primary Commodities (RF

Level of detail
Extended term

Term code
A02AZ

Term extended code
A02AZ#F02.A0EMD$F01.A04ZN$F27.A02AZ$F33.A0C39

Term name
Halibut, atlantic

Term extended name
Halibut, atlantic, PART-NATURE = Fish meat (as part-nature), S

Scope notes and links
Flat-fishes classified under the species Hippoglossus hippoglossus, commonly known as Atlantic halibut. The part consumed/analysed is not specified. When relevant, information on the part consumed/analysed has to be reported with additional facet descriptors. In case of data collections related to legislations, the default part consumed/analysed is the one defined in the applicable legislation.

en.wikipedia.org - google.co.uk

Implicit attributes:

Label	Value
Scientific n...	Hippoglossus hippogloss...
Common n...	Atlantic halibut
GEMSCode	WS4963
matrixCode	P1100000
foodexOld...	A.01.000897

A02AZ term code
unique five-character alphanumeric code

Scope Note description

Implicit attributes
- Scientific name,
- Common name,
- Matrix code *etc.*

FACETS

Facets (N=29): descriptors providing additional information for a particular aspect of a food

- Implicit facets (integrated in the catalogue)
- Explicit facets (added by users while coding a food/feed item)

Common examples

[F28] Process

[F04] Ingredient

[F09] Fortification agent

[F19] Packaging material

Facet groups

- [F02] Part-nature
- [F01] Source
- [F27] Source-commodities
- [F28] Process
- [F04] Ingredient
- [F06] Surrounding-medium
- [F08] Sweetening-agent
- [F09] Fortification-agent
- [F10] Qualitative-info
- [F17] Extent-of-cooking
- [F26] Generic-term
- [F21] Production-method
- [F18] Packaging-format
- [F19] Packaging-material
- [F03] Physical-state
- [F07] Fat-content
- [F11] Alcohol-content
- [F12] Dough-Mass
- [F20] Part-consumed-analysed
- [F22] Preparation-production-place
- [F23] Target-consumer
- [F24] Intended-use
- [F25] Risky-Ingredient
- [F29] Purpose-of-raising
- [F30] Reproductive-level
- [F31] Animal-age-class
- [F32] Gender
- [F33] Legislative-classes
- [F34] Host-sampled

FOODEX2 CODING

A flexible combination of base terms and facets

- Base term: **A03XF** is the code for

"Meat burger"

A03XF# F28.A0EJY

- Facet: **F28.A0EJY** means

"Grilling/griddling (high temperature cooking)"

A more complex, but typical case…

Base term followed (optionally) by '#' and a sequence of facet descriptors separated by '$'.

FOODEX2 CODING

Organic yoghurt, cow milk, semi skimmed, with cereals and raspberries

Facets

Corn flakes: **A00DD**
(Processed maize-based flakes)

Raspberries: **A01EP**
(Raspberries (red and yellow))

Instant oats: **A00DJ**
(Oat rolled grains, instant)

Organic production: **A07SE**
(Organic production)

Fat related qualitative information: **A077G**
(Semi-skimmed)

In a cup: **A07NT** *(Cup/pot)*
Made from glass: **A07PF** *(Glass)*

● **FoodEx2 Base term**

Yoghurt, cow milk, flavoured: A02NH
(Yoghurt, cow milk, flavoured)

FoodEx2 code string : A02NH#F04.A00DD$F04.A00DJ$F04.A01EP$F10.A077G$F21.A07SE$F18.A07NT$F19.A07PF

FoodEx2 training material:
https://www.efsa.europa.eu/en/data/data-standardisation

STEP FORWARD

Problems
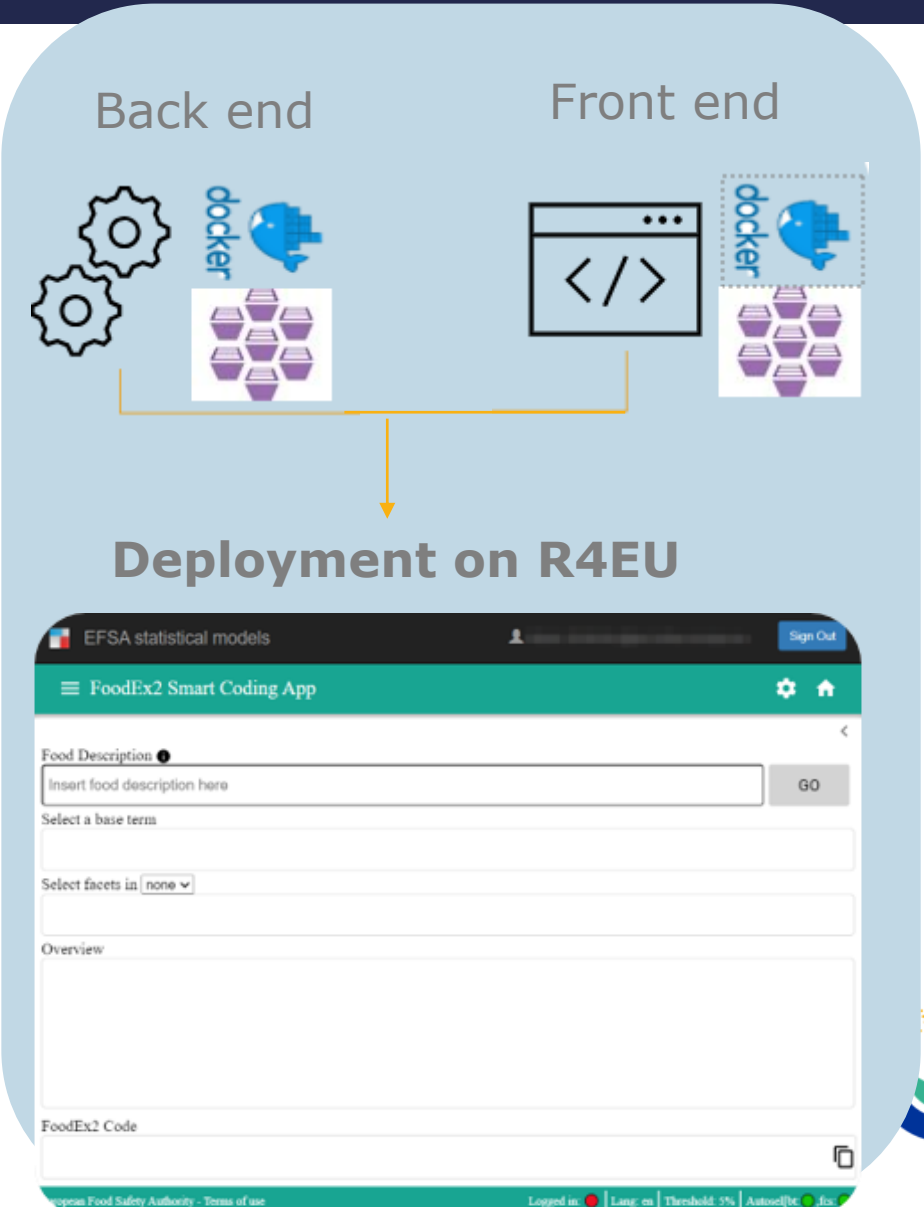
- FoodEx2 codification is **time-consuming**
- **Help data providers** during data submissions
- **Reduce the number of steps** for encoding
- **Reduce the knowledge required** for using the system
- **Improve the quality** of our data
- Make it available on the web **without local installations**

Solution

Artificial Intelligence ➔ FoodEx2 Smart Coding App

FOODEX2 SMART CODING APP
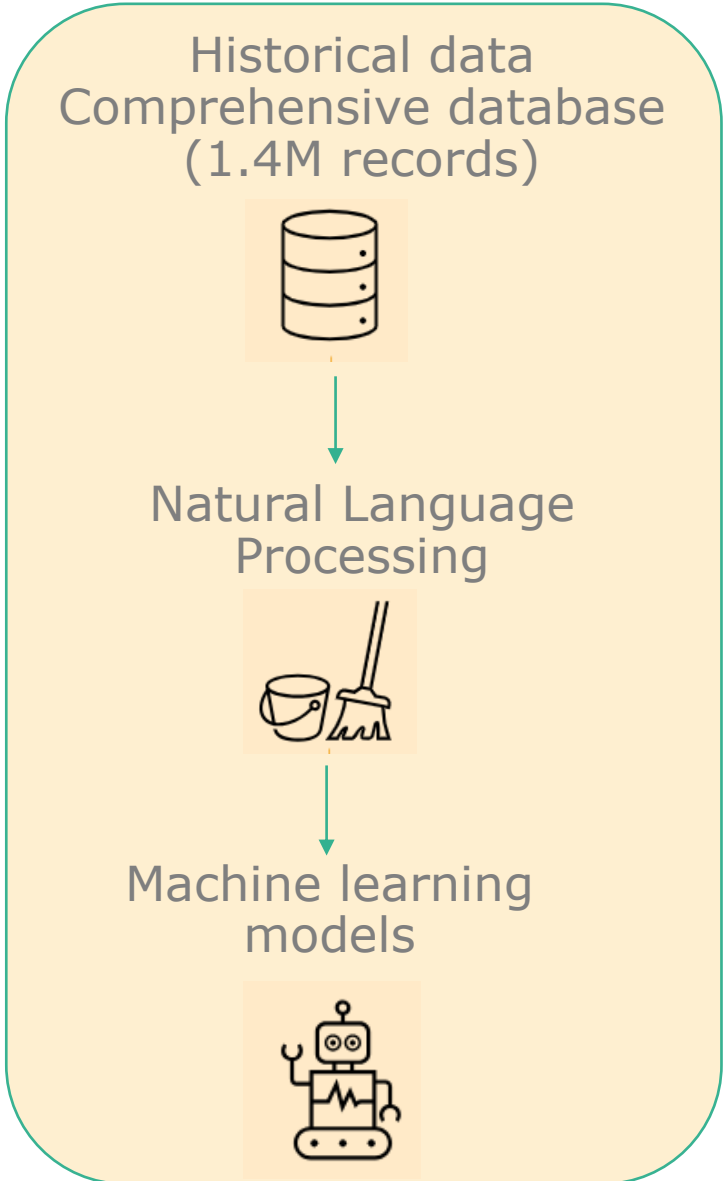
BACKEND – BASE TERM PREDICTION

Trained a multi class text classification model using SpaCy

predict

ENFOODNAME	BASETERM_CODE
OATS	A000F
Oat	A000F
OATS	A000F
Oat	A000G
BULGHUR WHEAT RAW	A000G
CEREALS,OATS,REG & QUICK,NOT FORT,DRY (USDA)	A000G
GRAIN, OAT, DRIED OR FLAKES, COOKING METHOD NOT APPLICABLE	A000G
GRAIN, OAT, FRESH, COOKING METHOD NOT APPLICABLE	A000G
Oat and similar-Oat grain	A000G
Oat grain	A000G
Oats (CY)	A000G
CEREALS,OATS,REG & QUICK,NOT FORT,DRY (USDA)	A000G
BARLEY STARCH	A000K

BACKEND – FACET CATEGORIES PREDICTION

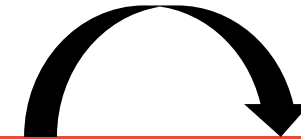Trained a multi label binary classification model using SpaCy

predict

ENFOODNAME	BASETERM_AND_EXPLICIT
POT BARLEY	A000P#F28.A07GG
cooked barley	A000P#F28.A07GG
cooked barley	A000P#F28.A07GG
PEARL BARLEY	A000P#F28.A07GH
BARLEY, WHOLE GRAIN, RAW	A000P#F28.A07HS
BARLEY WHOLE GRAIN DRIED	A000P#F28.A07KG
Grain, barley,cooked n.s.,dried/semi-dried n.s.,ambient department	A000P#F28.A07KG$F28.A0BA1
BARLEY COOKED	A000P#F28.A0BA1
Rice n.s.,stir fried/pan fried with fat,preservation method undefined,purchase department undefined	A001C#F26.A07XE$F28.A07GS
Rice n.s.,stir fried/pan fried with fat,dried/semi-dried n.s.,ambient department	A001C#F26.A07XE$F28.A07GS$F28. A07KG
Rice n.s.,stir fried/pan fried without fat,dried/semi-dried n.s.,ambient department	A001C#F26.A07XE$F28.A07GS$F28. A07KG

BACKEND – FACETS DESCRIPTOR PREDICTION

Trained several (for each facet category) multi
label classification models using SpaCy

predict

ENFOODNAME	BASETERM_AND_EXPLICIT
POT BARLEY	A000P#F28.**A07GG**
cooked barley	A000P#F28.**A07GG**
cooked barley	A000P#F28.**A07GG**
PEARL BARLEY	A000P#F28.**A07GH**
BARLEY, WHOLE GRAIN, RAW	A000P#F28.**A07HS**
BARLEY WHOLE GRAIN DRIED	A000P#F28.**A07KG**
Grain, barley,cooked n.s.,dried/semi-dried n.s.,ambient department	A000P#F28.**A07KG**$F28.**A0BA1**
BARLEY COOKED	A000P#F28.**A0BA1**
Rice n.s.,stir fried/pan fried with fat,preservation method undefined,purchase department undefined	A001C#F26.A07XE$F28.**A07GS**
Rice n.s.,stir fried/pan fried with fat,dried/semi-dried n.s.,ambient department	A001C#F26.A07XE$F28.**A07GS**$F28.**A07KG**
Rice n.s.,stir fried/pan fried without fat,dried/semi-dried n.s.,ambient department	A001C#F26.A07XE$F28.**A07GS**$F28.**A07KG**

BACKEND − PERFORMANCE

Average accuracy:

- Baseterm prediction: ~83%
- Facet categories prediction: ~87%
- Facets prediction (in category): ~87%

FOODEX2 SCA DEMO

1. **Insert free text description of the food you want to encode – Click 'Go'**
2. **Select base term from the list with top base terms correspoding to the free text introduced**
3. **Use drop down list to select facet category, and select facets from the list with top appropriate facets (multiple selections are possible)**

17

FOODEX2 SCA DEMO

4. **General overview field to confirm selections (in text)**
5. **FoodEx2 code generated live at each user interaction**

- Make it 'EFSA production IT': support from Service Desk (?)

- Resources (?)

Improvements wish list:

- Strengthen less frequent records

- Remove too frequent records

- Use a more complete dataset for training (different sources)

Next steps

Acknowledgements

Sofia Ioannidou (FoodEx2)
Carsten Behring (SCA)

anastasia.livaniou@efsa.europa.eu

REFERENCES

- The food classification and description system FoodEx 2 (revision 2)

https://efsa.onlinelibrary.wiley.com/doi/abs/10.2903/sp.efsa.2015.EN-804

- EFSA Catalogue Browser Download

https://github.com/openefsa/catalogue-browser/wiki

- EFSA Catalogue Browser User Guide

https://efsa.onlinelibrary.wiley.com/doi/10.2903/sp.efsa.2019.EN-1726

- EFSA FoodEx2 Smart Coding App (SCA)

https://r4eu.efsa.europa.eu/app/FoodEx2-SCA

- SCA User Guide (Version 1.0)

https://zenodo.org/doi/10.5281/zenodo.5596610

- SCA web component front-end

https://github.com/openefsa/foodex-web-component-frontend

STAY CONNECTED

SUBSCRIBE TO
efsa.europa.eu/en/news/newsletters
efsa.europa.eu/en/rss
Careers.efsa.europa.eu – job alerts

LISTEN TO OUR PODCAST
Science on the Menu –Spotify, Apple Podcast and YouTube

FOLLOW US ON TWITTER
@efsa_eu   @methods_efsa
@plants_efsa   @animals_efsa

FOLLOW US ON LINKEDIN
Linkedin.com/company/efsa

FOLLOW US ON INSTAGRAM
@one_healthenv_eu

CONTACT US
efsa.europa.eu/en/contact/askefsa

efsa
EUROPEAN FOOD SAFETY AUTHORITY