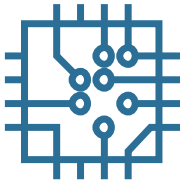
An application of advanced computer methods in Nutrition science

Assoc Prof Barbara Koroušić Seljak
Computer Systems department
Jožef Stefan Institute
Ljubljana, Slovenia



Computer Systems
Jožef Stefan Institute

Overview

- Few facts about my country & institute
- Cornerstones of dietary assessment & menu planning
- Problems & gaps
- On the way towards user-friendly solutions
- Current & future projects
- Q&A

Slovenia

By NuclearVacuum via Wikimedia Commons

- Inhabitants: 2 million (2,2 times more than at Tenerife)
- Area: 20,273 km^2 (10 times more than at Tenerife)
- Independence from Yugoslavia in 1991 (joined EU in 2004)
- Named as the 10^{th} best countries in Europe

Slovenian protected agricultural products and foodstuffs

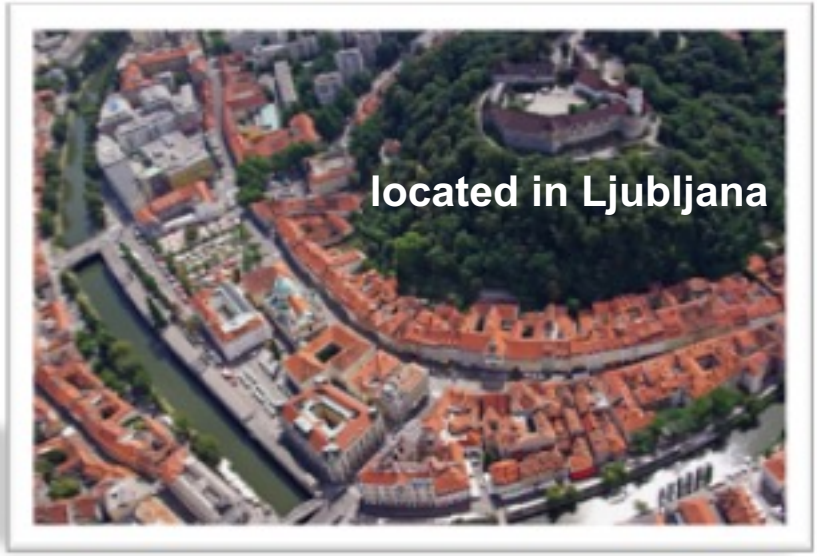
Slovenia's 24 distinct gastronomic regions are the result of the blending of local flavours with foreign influences

Slovenia's gross domestic product (GDP) stands at 91% of the EU-28 average.

Jožef Stefan Institute

The largest research institute in Slovenia

- Established in 1949
- Public institute

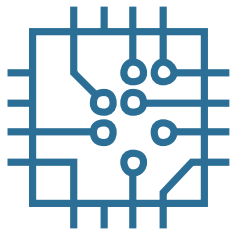


- More than 700 researchers in technology & Natural sciences
- 27 departments
- International Postgraduate School Jožef Stefan (Nanoscience&technology, ICT, Ecotechnologies)

Computer Systems department

http://cs.ijs.si

cs@ijs.si



Team:

14 researchers

- Electrical engineering & Computer science
- Mathematics & Physics

11 associates & students

- 4 PhD students
- 5 MSc students
- Visiting students (Marie-Curie fellowships)
- 1 dual doctorate student (University of Newcastle, AU)

Fields of research:

- Computational intelligence
  - Optimization with Meta-heuristic approaches
  - Modelling with Deep neural networks
- IoT / Embedded systems
  - Data sampling and collecting
- High-performance computing
  - Parallel computation
- Human-Computer Interactions
  - Serious games for telerehabilitation
  - Usability testing

Resent international projects

7x H2020 (PD-manager, REFRESH, RICHFIELDS, SYNERGY, UTOPIAE, TETRAMAX, SAAM)

- mHealth platform for Parkinson's disease management
- Reducing food waste
- Research infrastructure on consumer health and food intake
- Multi-objective optimisation with surrogate modelling and parallelization
- Uncertainty treatment and optimisation
- Customized low energy computing
- Supporting active ageing through multimodal coaching

1x ECSEL/ARTEMIS (MANTIS)

- Cyber-physical system based on proactive collaborative maintenance

1x FP7 ERA Chair (ISO-FOOD)

- Isotope techniques in food quality, safety and traceability

1x COST (ImAppNIO)

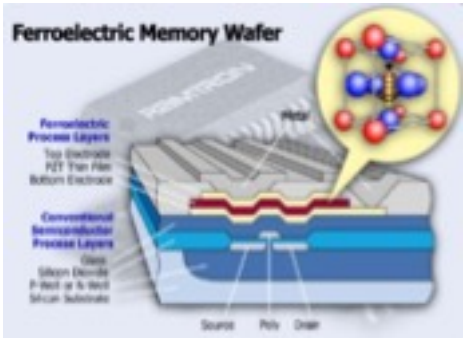
- Applicability of Nature-Inspired Optimisation

1x EFSA (EU Menu)

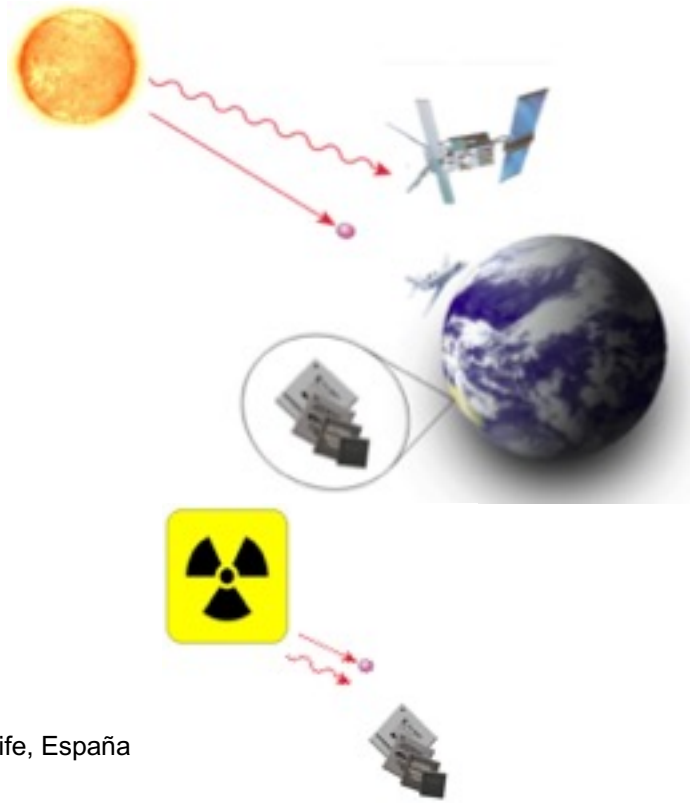
- Support to the national dietary survey 2017/19

Some results

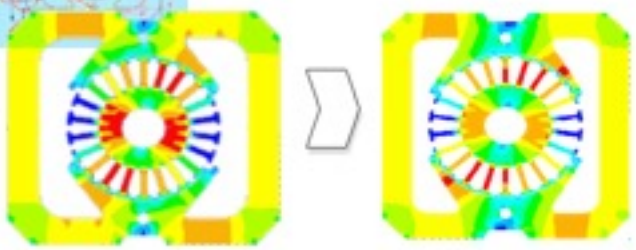
- Non-volatile solid-state data storage



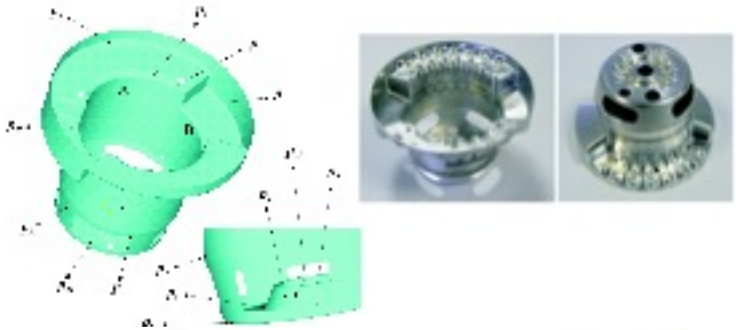
- Design of methods for on-line testing and system recovery
- Space radiation simulation in nuclear reactor



Some results (2)



- Efficient algorithms for massive-data processing
  - Biology-inspired algorithms (BIOMA)
  - Optimisation with advanced meta-heuristic approaches
  - Fast, effective
  - Simulation to optimize (stiffness, dimensions, stability, energy, cost...)
  - Use of finite-element methods (Ansys, Creo Parametric, SolidWorks, Catia, Inventor, SpaceClaim)



BIOMA

- Bioinspired Optimisation Methods and their Applications
- Biannual conference (April / May)
- Web site: https://bioma2018.sciencesconf.org

Relevant Twinning project

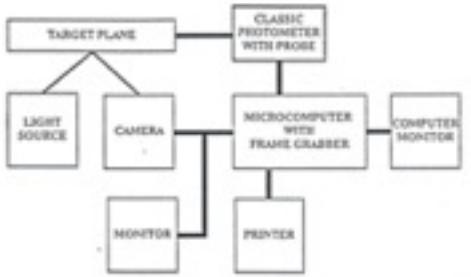
- SYNERGY – for Smart Multiobjective Optimisation
  - http://synergy-twinning.eu
  - IJS, USTL (parallelisation on large-scale heterogeneous architectures), CUAS (surrogate modelling and its deployment in optimisation problems).

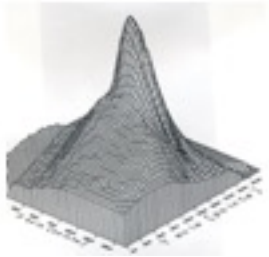
- Tea Tušar: https://dis.ijs.si/tea/publications.htm
  - Visualisation of Pareto fronts approximations in evolutionary multiobjective optimization

Some results (3)

- Simulation and optimization of mobile networks
  - GPU implementation of radio signal simulator for optimization of mobile

- Visual quality control of industrial parts
  - Visual quality control of insulating rubber parts in automotive industry



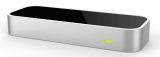


control of industrial robots – smart camera.

Some results (4)

- Usability testing

- 3D serious games
  - for Parkinson disease patients
  - Kinect (to detect arm motions)
  - Leap motion (to detect finger motions)



Dietary assessment

- Aimed to evaluate food intake of a given person
  - An assessment of anthropometrics and dietary intake
  - A physical examination and biochemical exams of blood/urine

- Measuring dietary data is important
  - To understand the role of diet in preventing chronic diseases
  - To uncover hospital malnutrition (in Slovenia around 40% of patients are at nutritional risk or malnourished)
  - To detect any other problem associated with diet
  - To advise on healthier nutrition (to plan healthier menus)

Cornerstones

- Measuring dietary (food consumption) data
  - *Person of age, sex, physical activity level*
  - *Consumed food or beverage or supplement of quantity at time*
  - *We are interested in the compositional value of the consumed food and drinks to be able to:*
    - *Educate about healthier diet*
    - *Design healthy menus*
- **Information** required for analysis of data:
  1. Dietary recommendations for this specific person (knowledge)
  2. Food composition data (data)

Methods & tools

- Required for dietary assessment
  I. To collect dietary (food consumption) data
  II. To create knowledge
  III. To collect food composition data
  IV. To analyze data (match food consumption with composition data)
  V. To design health menus

I. Dietary data collection

- Methods & tools
  - Food frequency questionnaire (FFQ) - frequency
  - Food diary – quantity
- Problems:
  - Motivation to complete a complex FFQ or keep a food diary for a longer period (more than 3-7 days)
  - Accuracy / memory
    - Taking pictures (DNN NutriNet)
    - Quantity / volume estimation (scale Libra, app with virtual reference object)
  - Merging data from FFQ and food diary
    - Improvement of the statistical Multiple Source Method (paper)

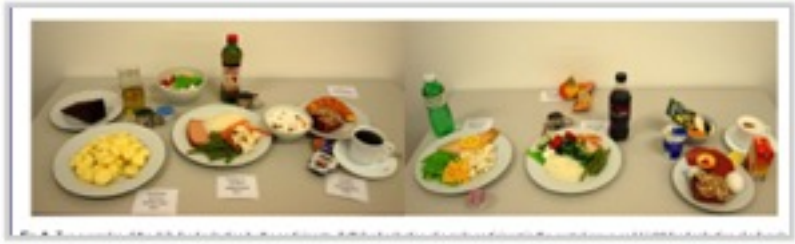
Food image recognition

Idea

1. Take a picture using a smartphone (or eButton)
2. Confirm detected & recognised food or drink
3. Confirm the estimated quantity
4. Train the system…

Approach

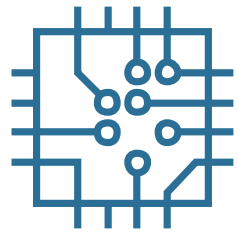
- Deep learning
- Use fake foods
  - To train the system
  - To learn about food quantity & volume



Beef Soup with Noodles

Creamy Vegetable Soup
Carrot Soup
Potato Goulash
Beef Soup with Noodles
Vegetable Soup

Deep learning



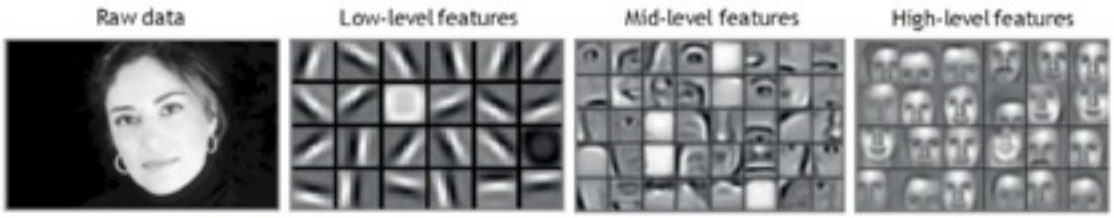
Vision problems

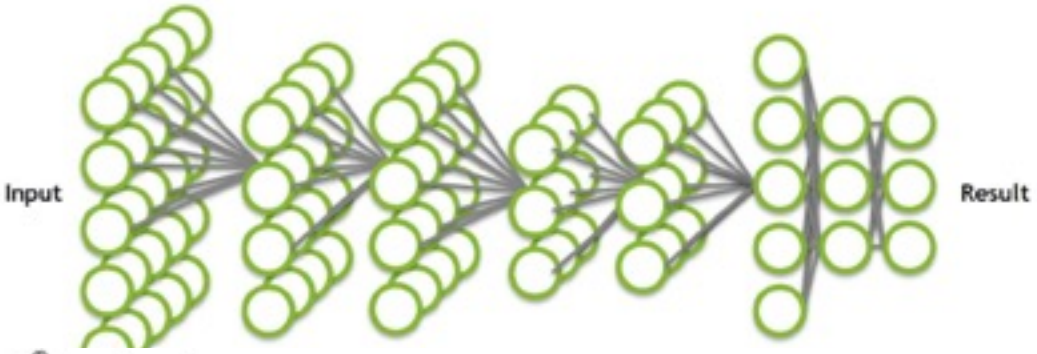
- Traditional approaches based on manually defined feature extractors (Hough transform, SIFT, SURF etc.)
  - Finding an optimal solution is nearly impossible

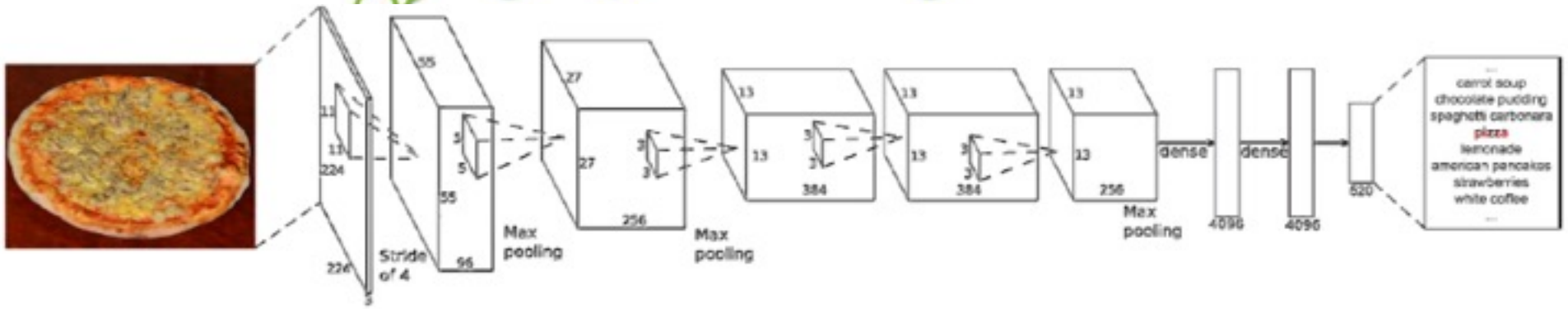
DL as a subfield of ML

- ANN with multiple hidden layers, which automatically learns about features from huge amounts of raw data
  - Successfully applied not only in computer vision, but also for speech recognition, machine translation
  - Robust method, enables parallelisation
  - Downside: black-box approach

Deep learning (2)







Images credit:

Jon Barker (NVIDIA)
March 2016

Simon Mezgec

Deep learning (3)

DL model

- In 2012, Alex Krizhevsky won the largest, most important Computer Vision challenge - ImageNet - with a Deep Learning model (AlexNet)
  - Since then, all winners of the ImageNet challenge use Deep Learning
  - Beating human accuracy in 2015
  - Game Go (AlphaGo) in 2016

Place for improvement

1. Create a dataset of food and drink images
2. DL model (NutriNet – based on AlexNet)
   - Extra convolutional layer
3. Training on GPU (NVIDIA TITAN X), 6-13x faster than on CPU

Deep learning (4)

1. Two extensive image datasets
   - **detection** dataset of our recipe and ImageNet images (130,517 images)
   - **recognition** dataset built using the Google Custom Search API (520 classes, 225,953 images)
   - 70% of images for training, 10% validation, 20% testing
2. These results are then fed into the detection DL model, which removes images that do not contain food or drink items.

Deep learning (5)

- NutriNet vs. AlexNet (on our datasets)
  - Faster training (14 hours)
  - Higher accuracy (recognition 86.72% / detection 94.47%)
    - Drink recognition
    - 520 food classes

Open problems:

How to recognise more foods on the plate?

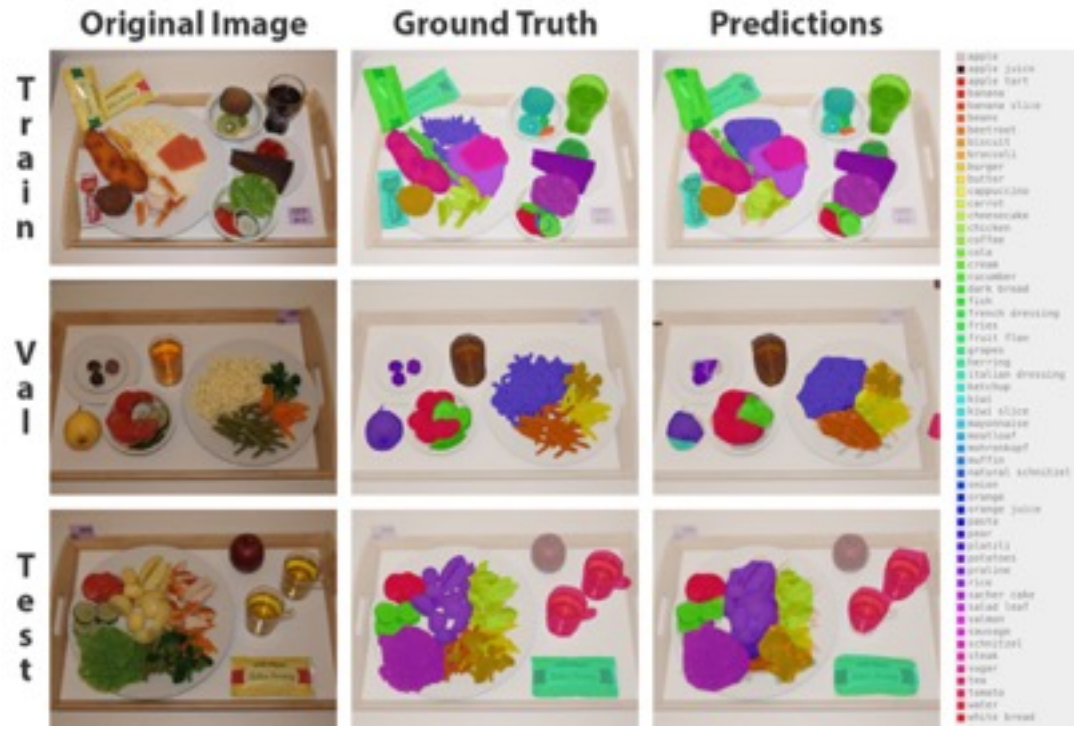
What kind of meat was used to prepare this dish?

Relevant paper

- S. Mezgec, T. Eftimov, T. Bucher, B. Koroušić Seljak, NutriNet: A Deep Learning Food and Drink Image Recognition System for Dietary Assessment, Nutrients – Open Access Human Nutrition Journal, Volume: 9, Issue: 7 (06/2017), 10.3390/nu9070657.

Deep learning (6)

- A single DL network to perform pixel-level segmentation and classification
- 124 study participants providing 55 food classes
  - 121 food images, in average 11 food items per plate
  - 500x375 pixels
- Training: 37 hours
- Pixel accuracy: 92.18%



Images credit:
Tamara Bucher & Simon Mezgec

Relevant publications

- S. Mezgec, B. Koroušić Seljak, Mixed deep learning and natural language processing method for fake-food image recognition and standardization to help automated dietary assessment, Public Health Nutrition (04/2018), 10.1017/S1368980018000708.
- T. Eftimov, S. Mezgec, P. Korošec, B. Koroušić Seljak, Mixed deep learning & natural language processing approach for food image detection, recognition and analysis aimed to estimate nutritional values. V: Abstract book, ISBNPA 2017, Annual Meeting International Society of Behavioral Nutrition and Physical Activity, June 5, 2017, Victoria, BC Canada. [S. l.]: Canadian Institutes of Health Research = CIHR, str.72; June 7 - 10, 2017, Victoria, BC, Canada.

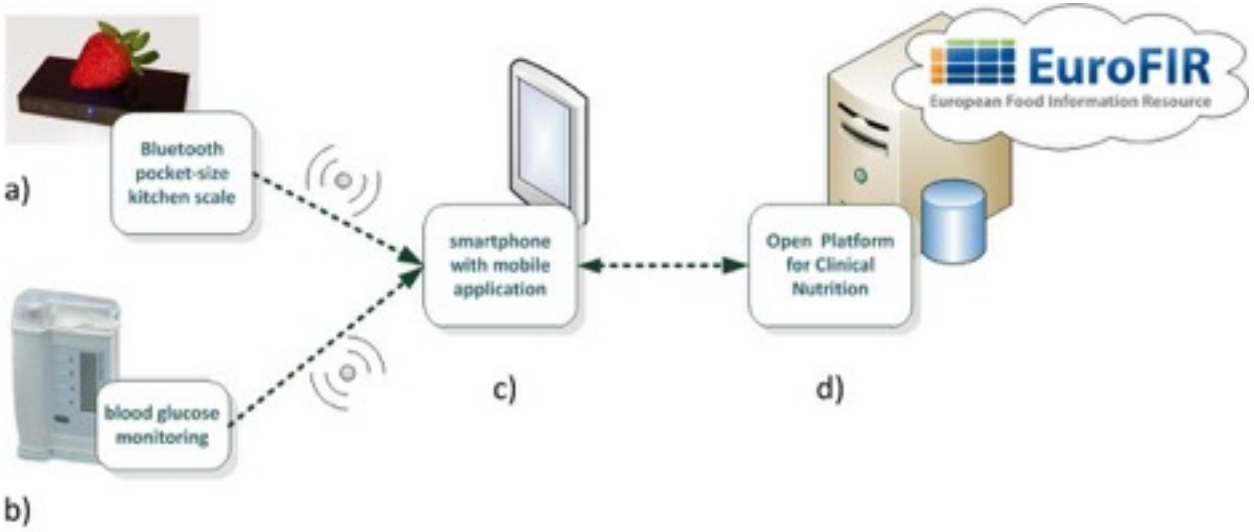
Wireless pocket-size kitchen scale

Idea

- Provide a diabetic patient with a light & cheap advice
  - ✓ Feasibility study
  - Clinical study

Approach

- Upgrade a kitchen scale with a Bluetooth module
  - Connect the scale with the Open Platform for Clinical Nutrition



Relevant paper

- G. Papa, B. Koroušić Seljak et al. Innovative pocket-size Bluetooth kitchen scale, Agro FOOD Ind Hi Tech (2018).

Related current projects (1)

- H2020 PD_manager
  - ✓ Mobile app & server solution
  - Open issue (training on fake food – volume estimation)

- H2020 SAAM
  - eButton

- The kitchen scale
  - ✓ Feasibility studies in collaboration with MDs & 6 patients
  - ✓ Patent
  - Open issue (clinical study)
  - New project Ballerina (Nutritics)

EFSA EU Menu
- Paper…

Methods & tools

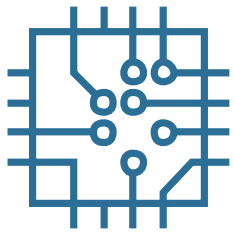
- Required for dietary assessment
  I. To collect dietary (food consumption) data
  II. To create knowledge
  III. To collect food composition data
  IV. To analyse data (match consumption with composition data)

II. Knowledge creation

- Methods and tools
  - Manual collection of evidence-based dietary recommendations and guidelines
  - Natural language processing (NLP) approaches

- Problems:
  - How to maintain up-to-date and complete knowledge?
    - drNER (NLP approach)
  - How to integrate knowledge?
    - Quisper ontology
  - How to validate automatically extracted knowledge by human experts? (Protégé, Stanford)

drNER – a rule-based Named-entity Recognition Method for Knowledge Extraction



- NER – automatic identification and classification of the words or phrases that describe important concepts (entities)
  - Terminological-driven NER methods (terminological resources)
  - Rule-based NER methods (regular expressions of dictionary information)
  - Corpus-based NER methods (annotated corpus and ML approaches)
- Information extraction in biomedical domain
  - Gene event extraction, cancer genetics, bacteria biotopes, seed development in plants
  - Protein-protein interactions, chemical/drug entity name recognition
  - Annotations of disease, phenotype, and averse reactions
  - New: Named-entity recognition in dietary domain (no annotated corpus!)
    - We do not use only dictionaries with concepts and synonyms (as terminological resources), but we allow the reuse of some corpus-based NERs that exist for some entities.

drNER (2)

- Entities: Food, Nutrient, Quantity/Unit, Action, Entity
  - S-Subject, P-Predicate, O-Object (location in the sentence)
  - Test corpus: 50 dietary recommendations summary from scientifically validated web sites and 50 abstracts of scientific publications

Babies need protein about 10 g a day.	Babies (S_1)	need (P_1)	-	protein (O_1)	10 g a day (O_1)
1 teaspoon of table salt contains 2300 mg of sodium.	-	contains (P_1)	table salt (S_1)	sodium (O_1)	1 teaspoon (S_1) 2300 mg (O_1)
Milk, cheese, yogurt and other dairy products are good sources of calcium and protein, plus many other vitamins and minerals.	-	are (P_1)	Milk, cheese, yogurt and other dietary products (S_1)	good sources of calcium and protein (O_1) many other vitamins and minerals (S_2)	-
Breast milk provides sufficient zinc, 2 mg/day for the first 4-6 months of life.	-	provides (P_1)	Breast milk (S_1)	sufficient zinc (O_1)	2 mg/day for the first 4-6 months of life (O_1)

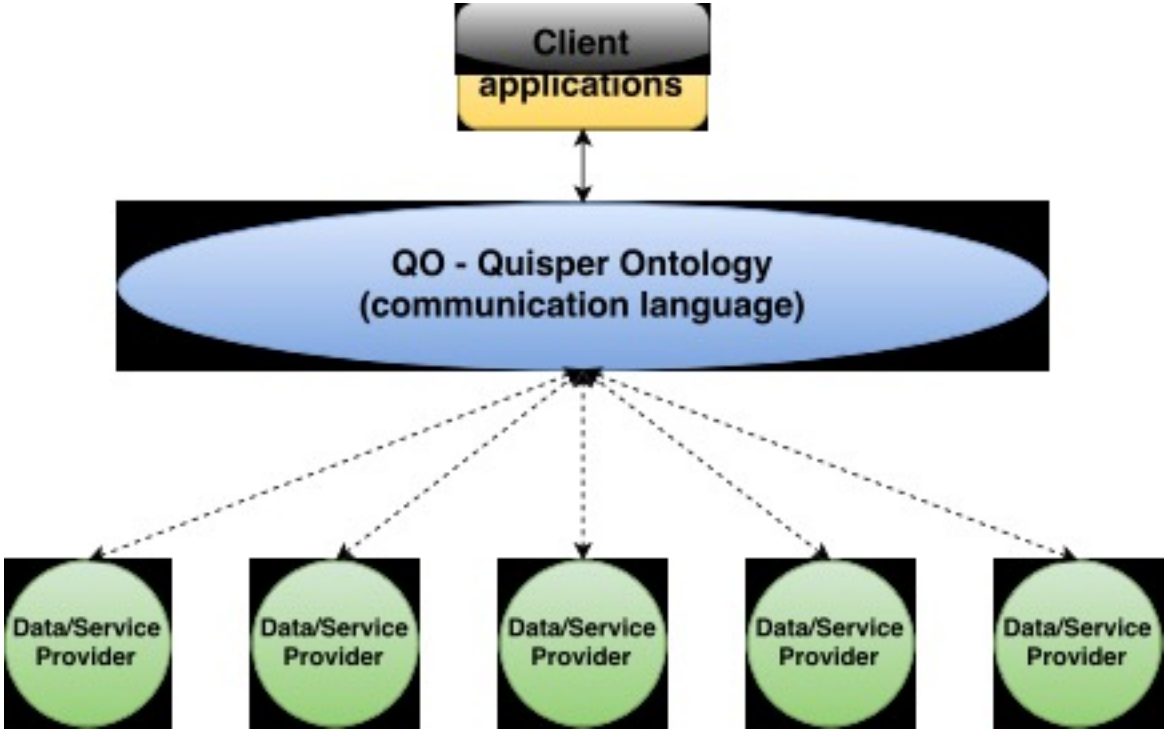
Related paper

- T. Eftimov, B. Koroušić Seljak, P. Korošec, A rule-based named-entity recognition method for knowledge extraction of evidence-based dietary recommendation, PLOS ONE, Volume: 12, Issue: 6 (06/2017), 10.1371/journal.pone.0179488.

Quisper & RICHFIELDS ontology

https://quisper.com

Quisper & RICHFIELDS ontology (2)



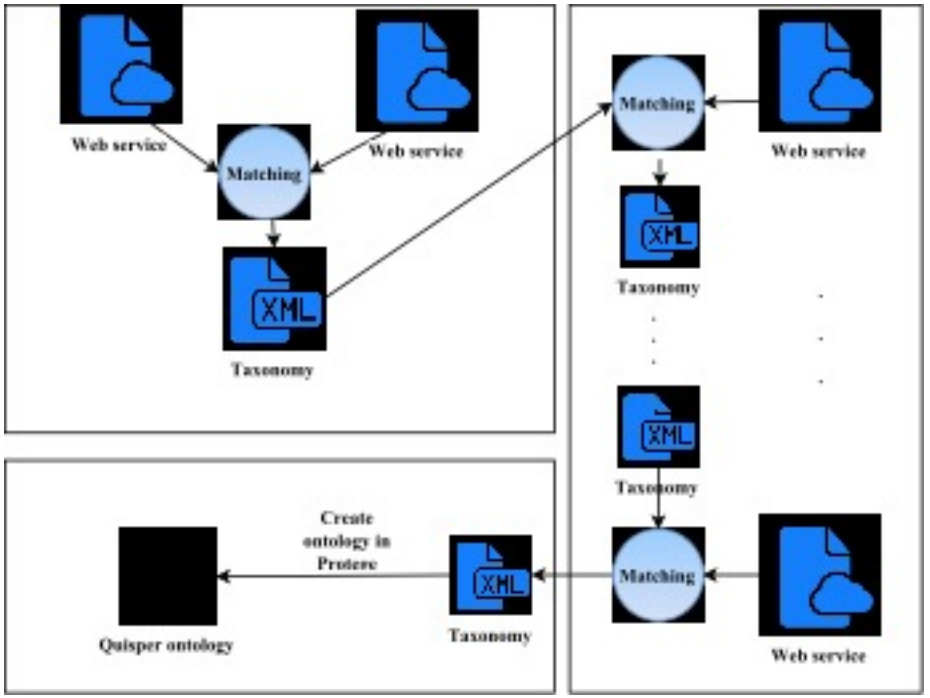
”carotene, alpha -” vs. ”alpha carotene”

“vitamin E alpha tocopherol eqiv from E vitamer actives” ***vs.*** *“Vitamin E (alpha-tocopherol)*

Quisper & RICHFIELDS ontology (3)

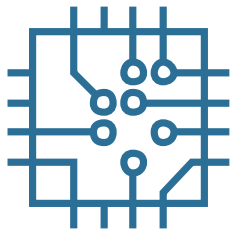
- POS tagging-probability weighted method

```
<Taxonomy>
  <WSproviders>
   <WSprovider name="EuroFIRFoodTransport"
   version="1.0.4"/>
   <WSprovider name="Food4me" version="1.0.0"/>
   <WSprovider name="SafeCape" version="1.0.0"/>
  </WSproviders>
  <Matching class="Component">
    <Descriptions>
       <Description source="EuroFIR"
       description="alpha-carotene,carotene,alpha-"
       code="CARTA"/>
       <Description source="Food4me"
       description="Alpha-carotene"
       code="4134833001"/>
    </Descriptions>
    <MatchScore>
       <MatchPair>
         <WSprovider name="EuroFIRFoodTransport"/>
         <WSprovider name="Food4me"/>
         <matchScore>0.1875</matchScore>
       </MatchPair>
     </MatchScore>
 </Matching>
       .
       .
       .
</Taxonomy>
```



RESTful web service

Quisper & RICHFIELDS ontology (4)



- (Boulos et al., 2015):
  - FoodWiki (Çelik, 2015)
  - AGROVOC (Caracciolo et al., 2012)
  - Open Food Facts (https://world.openfoodfacts.org/who-weare)
  - Food Product Ontology (Kolchin and Zamula, 2013)
  - FOODS (Diabetics Edition) (Snae and Brückner, 2008)

- The Quisper ontology covers a wider domain
- Not focused only on food-related data
- Easily *integrated* and *combined* with some of these 5 food ontologies

Related publications

- G. Ispirova, T. Eftimov, B. Koroušić Seljak, P. Korošec, Mapping Food Composition Data from Various Data Sources to a Domain-Specific Ontology, 9th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management, Volume: 2, KEOD (11/2017), 10.5220/0006504302030210; November 1 - 3, 2017, Funchal, Madeira, Portugal.
- T. Eftimov, G. Ispirova, P. Korošec, B. Koroušić Seljak, The RICHFIELDS Framework for Semantic Interoperability of Food Information across Heterogenous Information Systems, 10th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management, Volume: 1, KDIR (09/2018), 10.5220/0006951703150322, ISBN 978-989-758-330-8; September 18 - 20, 2018, Seville, Spain.

Related current projects (2)

- ERA Chair ISO-FOOD (http://isofood.eu)
  - Creating knowledge base on food safety and traceability
- H2020 RICHFIELDS (https://www.richfields.eu)
  - Entities on social determinants of food choice
- FP7 Quisper (ELIXIR Use case)
- H2020 SAAM (http://saamproject.eu)

Methods & tools

- Required for dietary assessment
  I. To collect dietary (food consumption) data
  II. To create knowledge
  III. To collect food composition data
  IV. To analyse data (match consumption with composition data)

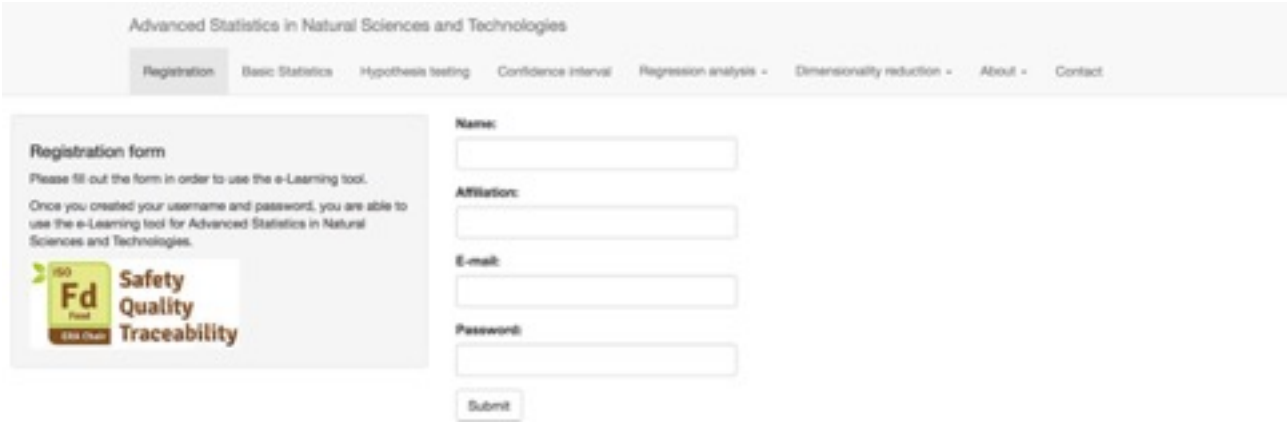
III. Food composition data collection

- Methods & tools
  - Food analysis
  - Food scanners (spectrometers Spectral Engines, SCiO and Tellspec)

- Problems
  - Statistical analysis (how to decide on the correct method? How to compare results of different methods?)
  - Always incomplete data
    - Data borrowing (MSc work – Gordana)

Deep statistics

- Chapter in book: Science within Food: Up-to-date Advances on Research and Educational Ideas
  - http://www.formatex.info/foodscience1/book/144-151.pdf
  - Web tool (statistical tutorial, methods for linear regression and principal component analysis (PCA) – can be extended to include other statistical methods).
    - The benefit: the e-learning tool checks for the required conditions of each statistical method and offers only the methods that are appropriate for analyzing the experimental data.



Related publication

- T. Eftimov, P. Korošec, B. Koroušić Seljak, Comparing multi-objective optimization algorithms using an ensemble of quality indicators with deep statistical comparison approach. In Computational Intelligence (SSCI), 2017 IEEE Symposium Series on (pp. 1-8) (12/2017). IEEE; November 27 - December 1, 2017, Honolulu, Hawaii, USA.

Borrowing missing data

- Which data are popular today?

  675 ingredients collected from

  - 800 recipes for main dishes
  - 200 recipes for desserts

  Approx. 40,000 foods from FCDBs

Borrowing missing data (2)

- UK

- Slovenia

- Danmark

Borrowing missing data (3)

- The main problem:
  - How to select data from other FCDBs that are of high quality and belong to the same or highly comparative food item?
  - By using a statistical approach based on null-hypothesis testing
    - using the interpretations H0: There is no effect/difference between the separate data sets / H1: There is an effect/difference between the separate data sets

Borrowing missing data (4)

- MIGHT: Missing Data Imputation using Null-Hypothesis Testing (results in order of sec)
  - in 69.2% of the cases the methodology gives more accurate results than other state-of-the-art approaches used for missing value imputation.
  - Submitted paper

Related publication

- G. Ispirova, Quality improvement of food composition databases by using methods from natural language processing and statistics, MSc Thesis, 2018

Related current projects (3)

- H2020 RICHFIELDS
- H2020 SAAM
- National ‘Food switch’ app
  - Enrich data from the food label with compositional data from generic foods

Methods & tools

- Required for dietary assessment
  I. To collect dietary (food consumption) data
  II. To create knowledge
  III. To collect food composition data
  IV. To analyse data (match consumption with composition data)

IV. Dietary data analysis

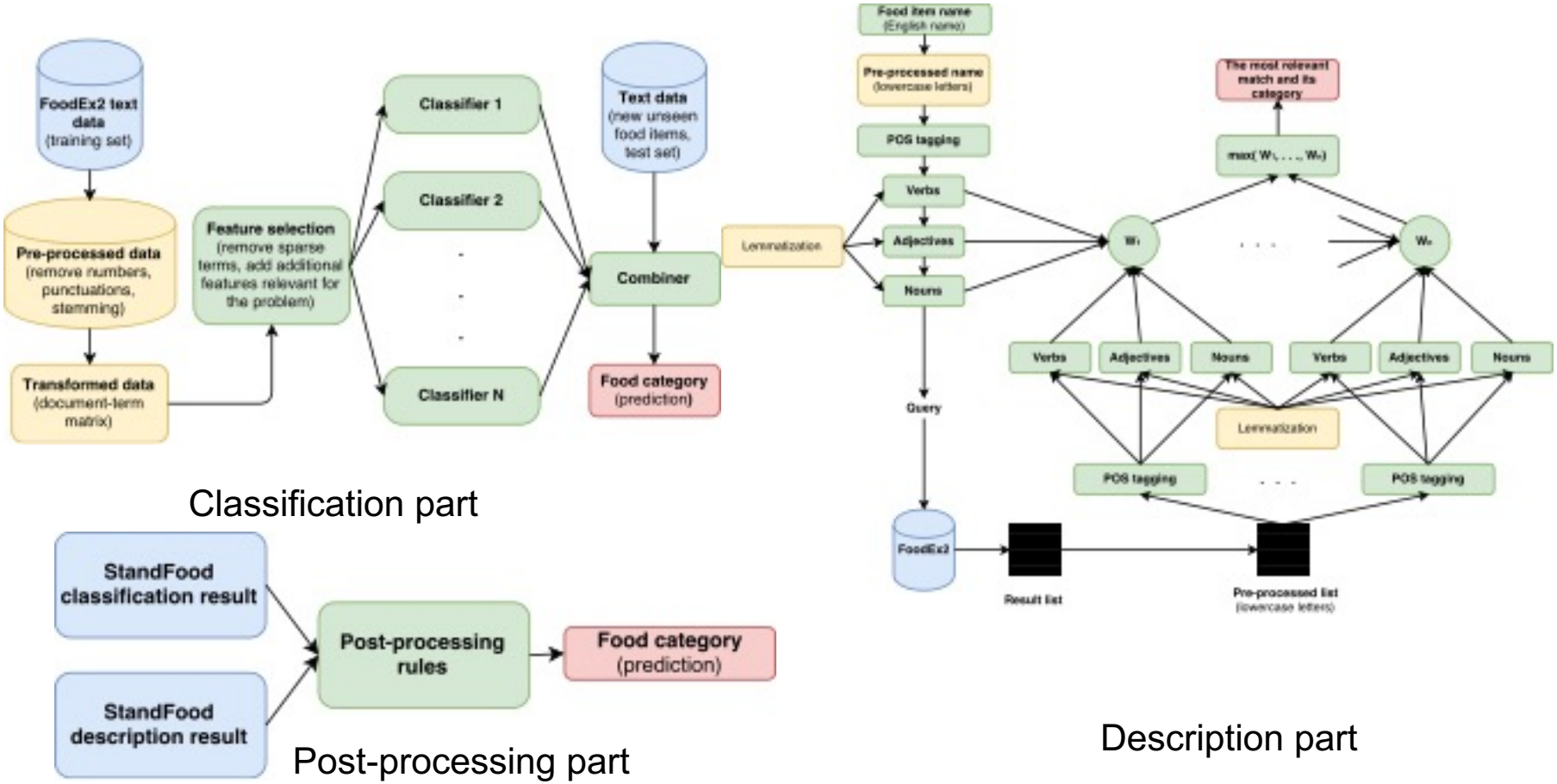
- Methods & tools
  - Food matching (consumption to composition data)

- Problems
  - Food consumption and composition data are classified and described using different systems (FoodEx2, GloboDiet etc. / LanguaL)
    - StandFood (FoodEx2) / FoodEx2 to LanguaL
  - Missing information about recipes
    - Recipe calculation
  - Validation by human experts

Food matching

- Novel method: StandFood / Standardization of Foods (using food names)
  - Example on EFSA FoodEx2 food classification & description system:
    - Four food categories: raw (r), derivatives (d), simple (s) and aggregated (c) composite foods
    - FoodEx2 code example (A03BG#F09.A0EXH$F10.A077L$F21.A07SE)
      - *nectar, orange (A03BG),*
      - *FORTIFICATION AGENT = 53 Calcium (F09.A0EXH)*
      - *QUALITATIVE INFO = Sugar free (F10.A077L)*
      - *PRODUCTION METHOD = 54 Organic production (F21.A07SE)*

StandFood classification, description, and post-processing part



StandFood results

- 532 instances from Slovenian dataset
- 89% accuracy of the **classification** part and 79% accuracy of the **description** part

Food Item	StandFood FoodEx2 Code	StandFood Relevant FoodEx2 Item	Manual FoodEx2 Code
Mushroom soup	A041R	Mushroom soup	A041R
Prepared green salad	A042C	Mixed green salad	A042C
Meat burger	A03XF	Meat burger no sandwich	A03XF
Yeast	A049A	Baking yeast	A049A
Brown sauce (gravy, lyonnais sauce)	A043Z	Continental European brown cooked sauce gravy	A043Z
Cow milk, <1% fat (skimmed milk)	A02MA	Cow milk skimmed low fat	A02MA
Supplements containing special fatty acids (e.g., omega-3, essential fatty acids)	A03SX	Formulations containing special fatty acids (e.g., omega-3 essential fatty acids)	A03SX
Durum wheat flour (semola)	A004C	Wheat flour durum	A004F
Gingerbread	A00CT	Gingerbread	A009Q$F14.A07GX
Cherry, fresh	A01GG	Cherries and similar	A01GK
	A01GH	Sour cherries	
	A01GK	Cherries sweet	
	A0DVN	Nanking cherries	
	A0DVP	Cornelian cherries	
	A0DVR	Black cherries	

Related publications

- T. Eftimov, P. Korošec, B. Koroušić Seljak, StandFood: Standardization of Foods Using a Semi-Automatic System for Classifying and Describing Foods According to FoodEx2, Nutrients - Open Access Human Nutrition Journal, Volume: 9, Issue: 6 (05/2017), 10.3390/nu9060542.
- T. Eftimov, G. Ispirova, B. Koroušić Seljak, Peter Korošec, A semi-automatic system for classifying and describing foods according to FoodEx2, 3rd IMEKOFOODS Metrology Promoting Harmonization& Standardization in Food & Nutrition, (10/2017); October 1 - 4, 2017, Thessaloniki, Greece.

Recipe calculation

- POS tagging probability weighted method for matching the Internet recipe ingredients with food composition data
  - Data – collection of 721 recipes written in English (1,615 different names of ingredients)
  - EuroFIR FCDB (44,033 English names of food items)
  - Data preprocessing (remove punctuations, convert each name in lower case letter, whitespace tokenization, lemmatization, only for nouns; manually created rules (without skin; skinless)

Recipe calculation (2)

- Perfect - 74.92%
- Very similar - 16.90 %
- Similar - 4.84%
- Incorrect - 3.34%

Related paper

- B. Koroušić Seljak et al, Identification of Requirements for Computer-Supported Matching of Food Consumption Data with Food Composition Data, Nutrients - Open Access Human Nutrition Journal, 2018 Mar 30;10(4). pii: E433, 10.3390/nu10040433.

Related current projects (4)

- FP7 QuaLiFY (http://www.qualify-fp7.eu)
- H2020 RICHFIELDS (http://www.richfields.eu)
- EuroFIR (new tool for food matching)
- H2020 SAAM (https://saam2020.eu)

BIG thank you

- Tome Eftimov
- Simon Mezgec
- Gordana Ispirova
- Peter Korošec
- Drago Torkar
- Eva Valenčič

- Find more about us from ResearchGate
- Contact: barbara.korousic@ijs.si