POST-probability weighted method for matching the Internet recipe ingredients with food composition data

Tome Eftimov [1,2], Barbara Koroušić Seljak [1]

[1] Computer Systems Department, Jožef Stefan Institute, Jamova cesta 39, 1000 Ljubljana, Slovenia

[2] Jožef Stefan International Postgraduate School, Jamova cesta 39, 1000 Ljubljana, Slovenia

Motivation

- Nowadays, one of the major causes of global health problems is strongly related to **poor** and **unbalanced diet**.
- A lot of information about **healthier diet** is presented in different forms, available in books, magazines, television and Internet.
- People are lacking of knowledge about all **nutrition principles** and also lack of time and motivation to explore the resources where this kind of information is presented.
- **Food composition databases (FCDBs)** provide detailed information on nutritional composition of food.
- A new method for **matching the recipe ingredients** extracted from Internet to nutritional data from FCDBs.

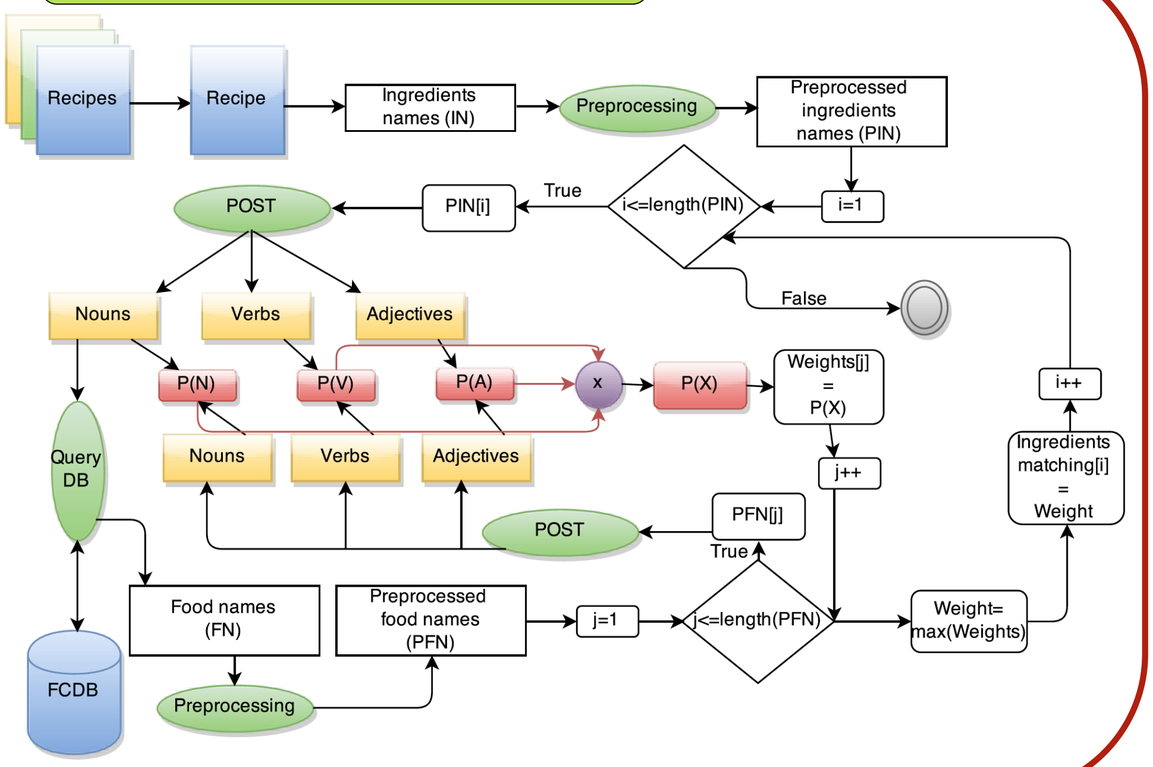
Problem definition

- Ingredients –food composition data matching
- People use human language to write the names of the used ingredients (unstructured form)
- “salt, iodised”, “iodised salt”, “salt-iodised”
- Ingredient synonymy problem
- Preparation method of the ingredient – different nutritional principles

Proposed method

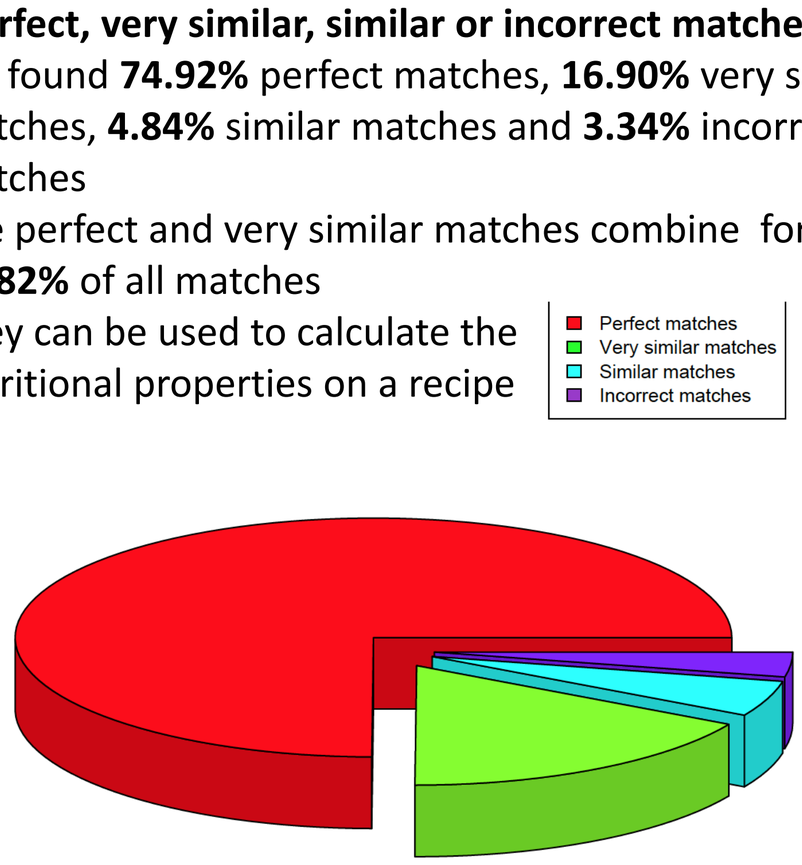
- Probability method with witch we assigned a weight on each matching
- The match with the highest weight is the most relevant one
- POST to capture the information of the names of the ingredients and the names of the food analyses from FCDB
- The weight we assign is the similarity between the ingredient name and the name of the food analysis from the FCDB
- **P(X) = P(N) P(V) P(A)**, where P(N), P(V), and P(A) are the similarities between the sets of nouns, verbs, and adjectives, respectively
- $P(Y) = \frac{J(Y_1,Y_2)+\frac{1}{|Y_1 \cup Y_2|}}{1+\frac{2}{|Y_1 \cup Y_2|}}$, where Yϵ {N, V, A}, and it is a combination of Jaccard index and Laplace probability estimate

Flowchart of the method



Experiment results

- The data we used is a collection of 721 recipes written in English, from which we extracted 1,615 different names of ingredients
- We used 44,033 English names of food analyses from EuroFIR database (http://www.eurofir.org/)
- Each match can be in one of the four categories (**perfect, very similar, similar or incorrect matches**)
- We found **74.92%** perfect matches, **16.90%** very similar matches, **4.84%** similar matches and **3.34%** incorrect matches
- The perfect and very similar matches combine for **91.82%** of all matches
- They can be used to calculate the nutritional properties on a recipe



Conclusion and future work

- New ingredients – food composition data matching method
- The weighted data can be used to help more other models, which can be obtained using machine learning approaches
- Compare recipe nutritional values calculated by our method with the values acquired by chemical analysis of dishes prepared in reference to the recipes

This work was supported via the project ISO-FOOD, which received funding from the European Union’s Seventh Framework Programme for research, technological development and demonstration under grant agreement no 621329 (2014-2019)