Matching foods from EuroFIR databases to FoodEx by using a semi-automatic system for classifying and describing foods

Tome Eftimov1,2, Peter Korošec1,3, Barbara Korušić Seljak1, Mark Roe4,
Aida Turrini5, Paul Finglas6

1 Computer Systems Department, Jožef Stefan Institute, Slovenia
2 Jožef Stefan International Postgraduate School, Slovenia
3 Faculty of Mathematics, Natural Sciences and Information Technologies, Slovenia
4 Quadram Institute Bioscience, UK
5 Crea-Council for Agricultural Research and Economics - Research Center for Food and Nutrition, Italy
6 EuroFIR AISBL, Brussels, Belgium

INTRODUCTION

• FoodEx2 – a standardized food classification and description system provided by EFSA.
• FoodEx2 uses facets to describe food properties and aspects from various perspectives making it easier to compare food consumption data from different sources and perform more detailed data analysis.
• It consists of three types of food categories representing three different levels in the food chain, moving from raw commodities (r), to derivatives (d) and composite foods (simple (s) and aggregated (c)).
• It also includes two types of terms: list terms and facet descriptors. List terms are represented by a code and describe food items. Facet descriptors are elements of additional information added to the list terms, such as treatments received, production method, fat content, and qualitative information.

OBJECTIVE

• As most food-related datasets rely on different standards, which are related to the same concepts but use different terminology and classification, we decided to solve the problem of data harmonization by linking the datasets to FoodEx2 data.
• A lot of food-related datasets are lacking in FoodEx2 codes because the process of classification and description has to be manually performed – a process that is laborious and requires good knowledge of the system.
• To match foods from EuroFIR databases to FoodEx2 data, we used a semi-automatic system for classifying and describing foods according to FoodEx2 called StandFood [1].

METODOLOGY

StandFood consists of three parts:
• Classification;
• Description;
• Post-processing rules.

MAIN FINDINGS

StandFood Classification part

StandFood Description part

StandFood Post-processing part

CONCLUSION

• An evaluation of the system was made using datasets of the food composition databases from Slovenia, Greece, Italy, and the United Kingdom. StandFood obtained an accuracy of 89% for the classification part and 79% for the description part.
• The system can be used to find FoodEx2 codes for food-related datasets.
• According to the time needed to code 682 foods, only few minutes (cca. five minutes) are required because the current version of the system is programmed in a sequential way. Parallel programming would allow the same task to be carried in seconds.