July 2020

An Insight into Food Semantics: Review, Analysis, and Lessons Learnt over Food-related Studies

Gorjan Popovski [a,b], Gordana Ispirova [a,b], Eva Valenčič [a,b], Riste Stojanov [c], Tome Eftimov [a], Barbara Koroušić Seljak [a,1]

[a] *Computer Systems Department, Jožef Stefan Institute, 1000 Ljubljana, Slovenia*
[b] *Jožef Stefan International Postgraduate School, 1000 Ljubljana, Slovenia*
[c] *Faculty of Computer Science and Engineering, Ss. Cyril and Methodius, University, 1000 Skopje, North Macedonia*

A great amount of work has been done in predictive modelling in the past decades. This is made possible by the existence of biomedical vocabularies and standards. Despite the availability of such resources in the health domain, the domain of food and nutrition is low-resourced and can greatly benefit from such methods. Lancet Planetary Health published that starting from 2019 the focus will be on the links between food systems, human health, and the environment. Several food ontologies exist, but each developed for a specific application scenario. Hence, in 2019, the Big Food and Nutrition Data Management and Analysis (BFNDMA) workshop started at the IEEE International Conference on Big Data [2], focusing focuses on methodologies for big data management and analysis for food and nutrition data.

Recently, in an effort to tackle the task of Information Extraction (IE) from unstructured text, several methodologies were proposed. DrNer [1] is a rule-based NER method for extracting information from evidence-based dietary recommendations. The authors have extended the methodology by creating a novel food NER method named FoodIE [2], which solely focuses on extracting food entities. It incorporates computational linguistic rules and semantic information. The authors have compared it to other existing food NER methods, showing that FoodIE provides the most promising results [3]. The collection of such information from different data sources is represented in various unstandardized ways, leading to the task of data normalization. It is a crucial task to facilitate and enable further analyses. Hence, StandFood [4] has been introduced - a semi-automatic system for classifying and describing foods according to FoodEx2. It is based on lexical similarity between food names. In a recent paper [5], the authors have conducted a domain-coverage analysis of an existing language for describing foods (LanguaL) by using Representation Learning (RL) methods, finding that the coverage of the food domain does not link the concepts together well, accenting the need for future efforts in food data normalization. FoodBase [6] is the first data corpus consisting of recipes (total of 23,000) annotated with the food entities found in them. The authors extend this work by propos-

[1] Corresponding E-mail: barbara.korousic@ijs.si.
[2] http://cs.ijs.si/bfndma/BFNDMA.html

ing a food data normalization method (FoodOntoMap [7]) that performs food concept mapping across different food semantic resources. It is based on the use of food NER methods to perform the mapping. In addition, a visualization tool (FoodViz [8]) provides a framework aimed at making the links between different food standards understandable by food subject-matter experts.

Finally, the authors in a recent paper [9] have performed a non-English language (i.e. Slovenian) case study where they employ RL methods in order to address the semantic similarity between food products' names.

Addressing open gaps, the extracted food entities can further be linked with entities from other domains (e.g. health, biomedicine, consumer and social sciences). This can help in reducing knowledge gaps that inhibit public health goals as well as the optimal development of scientific, agricultural and industrial policies, which requires relevant information from all food science domains, such as food safety, food authenticity and traceability, food sustainability, etc. Coupled with Representation Learning techniques, this can pave the way for methods to extract information in order to improve personalized nutrition and medicine, as well as public health.

Acknowledgments. This research was supported by the Slovenian Research Agency (research core grant number P2-0098), and the European Union's Horizon 2020 research and innovation programme (FNS-Cloud, Food Nutrition Security) (grant agreement 863059). The information and the views set out in this publication are those of the authors and do not necessarily reflect the official opinion of the European Union. Neither the European Union institutions and bodies nor any person acting on their behalf may be held responsible for the use that may be made of the information contained herein.

References

[1] T. Eftimov, B. Koroušić Seljak, and P. Korošec, "A rule-based named-entity recognition method for knowledge extraction of evidence-based dietary recommendations," *PloS One*, vol. 12, no. 6, p. e0179488, 2017.

[2] G. Popovski, S. Kochev, B. Koroušić Seljak, and T. Eftimov, "Foodie: A rule-based named-entity recognition method for food information extraction," in *Proceedings of the 8th International Conference on Pattern Recognition Applications and Methods, (ICPRAM 2019)*, pp. 915–922, 2019.

[3] G. Popovski, B. Koroušić Seljak, and T. Eftimov, "A survey of named-entity recognition methods for food information extraction," *IEEE Access*, vol. 8, pp. 31586–31594, 2020.

[4] T. Eftimov, P. Korošec, and B. Koroušić Seljak, "Standfood: Standardization of foods using a semi-automatic system for classifying and describing foods according to FoodEx2," *Nutrients*, vol. 9, no. 6, p. 542, 2017.

[5] G. Popovski, B. Paudel, T. Eftimov, and B. Koroušić Seljak, "Exploring a standardized language for describing foods using embedding techniques," in *2019 IEEE International Conference on Big Data (Big Data)*, pp. 5172–5176, 2019.

[6] G. Popovski, B. Koroušić Seljak, and T. Eftimov, "FoodBase corpus: a new resource of annotated food entities," *Database*, vol. 2019, 11 2019. baz121.

[7] G. Popovski., B. K. Seljak., and T. Eftimov., "Foodontomap: Linking food concepts across different food ontologies," in *Proceedings of the 11th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management - Volume 2: KEOD,*, pp. 195–202, INSTICC, SciTePress, 2019.

[8] R. Stojanov, G. Popovski, N. Jofce, D. Trajanov, B. Koroušić Seljak, and T. Eftimov, "Foodviz: Visualization of food entities linked across different standards," in *Proceedings of The Sixth International Conference on Machine Learning*, p. In Press., 2020.

[9] G. Popovski., G. Ispirova., N. Hadzi-Kotarova., E. Valenčič., T. Eftimov., and B. K. Seljak., "Food data integration by using heuristics based on lexical and semantic similarities," in *Proceedings of the 13th International Joint Conference on Biomedical Engineering Systems and Technologies - Volume 5: HEALTHINF,*, pp. 208–216, INSTICC, SciTePress, 2020.