

Managing Incompleteness and Validating the Content of Nutritional Food Sources using FoodOn as Pivot Ontology

Patrice BUCHE^{a, b, 1}, Julien CUF^b, Stéphane DERVAUX^c, Liliana IBANESCU^c, Alrick OUDOT^b and Magalie WEBER^d

^a*LIRMM, Univ Montpellier, CNRS, INRIA GraphIK, Montpellier, France*

^b*IATE, Univ Montpellier, INRAE, CIRAD, Montpellier SupAgro, Montpellier, France*

^c*UMR MIA-Paris, AgroParisTech, INRAE, University Paris-Saclay, Paris, France*

^d*BIA INRAE, Nantes, France*

Abstract. In order to correctly assess the nutritional quality of a raw or manufactured food product, the first step is to obtain the associated nutritional values. Food composition databases (FCDBs) managed at national level provide values for nutrients of foods. Unfortunately, values associated with some nutrients of interest may be lacking in the FCDB of the country in which the nutritional quality must be assessed and finding values associated with nutrients for similar foods in other FCDBs is a way to deal with incompleteness. An additional issue arises because the vocabulary used to denote a given food in a given FCDB is usually different from the one used in others. In this paper, the authors address the problem of retrieving the nutritional value of foods by querying different FCDBs through FoodOn used as pivot ontology. The article presents a new food source alignment method between two FCDBs. The method has been evaluated on the French and United States food nutritional FCDBs. The proposed solution for the incompleteness management task has been assessed with a real use case.

Keywords. Ontology alignment, Food composition databases, FoodOn, LanguaL

For national and international food trading, it is a challenge to automatically generate the nutrition information panel required by regulation for raw or manufactured food products in many countries. The first challenge is to identify the nutritional values of the raw or manufactured food product either by its identification in an appropriate Food Composition DataBase (FCDB) [1], or by designing a specific experimental analysis procedure which may require high expertise, performant analysis tools and time. Unfortunately, values associated with nutrients of interest for a food may be lacking in the FCDB of the country in which the nutritional quality must be assessed. Finding values associated with nutrients for similar foods in other FCDBs is a way commonly used by nutritionists to deal with incompleteness.

¹ Corresponding Author

This paper addresses the problem of semi-automatically identifying the nutritional value of raw or manufactured food products by querying different Food Composition Databases through a pivot vocabulary (named a master-code approach in [2]) in order to deal with the lack of nutrient values. An additional issue arises because the vocabulary used to describe the ingredients of a food or a recipe in a given FCDB is usually different from the vocabulary used in others.

A lot of efforts have been done during the 30 last years in order to harmonize food nutritional data sources through world wide networks like INFOODS [6] or EUROFIR [7]. A lot of standards exist concerning food classification and description systems, as reviewed and compared in [2]. LanguaL [5], a multilingual thesaurus using faceted classification, is used in major food composition databases, *e.g.* in the United States [9], Europe [7] and France [8] to define a food item by a set of standard controlled terms. Moreover, FoodOn [3] is an ontology, initially based on a conversion of the LanguaL thesaurus, integrated with other resources and aiming to be the open standard controlled vocabulary for food science. In order to be able to integrate these major FCDBs and terminologies, called food sources, this paper proposes to align on FoodOn (that is therefore used as a pivot) a given food using both its LanguaL and English terminological descriptions commonly available in all FCDBs. In this proposed scenario, two foods from two different food sources being indexed with the same LanguaL description and same terminological English description are assumed to represent the same food.

This flash presentation presents a new method to align a food source on a target one (i.e. FoodOn) using both food LanguaL description and the English terminological description. Our method has two main steps: (1) transformation of the food sources in food ontologies and (2) food product alignment computation based on semantic and syntactic information. In this approach, aligning a new FCDB on FoodOn will take benefit of FCDBs already aligned on FoodOn. Indeed, it allows by transitivity an automatic alignment of the new FCDB on FCDBs already aligned and avoids bilateral alignment efforts between FCDBs. During the French national Meatylab project gathering industrial and academic partners, this approach has been implemented in a new application called MultiDB explorer which currently integrates several national FCDBs including Ciqua and USDA. MultiDB explorer has been in particular used to deal with the lack of values in Ciqua for 3 nutrients of interest selected by industrial partners (Vitamin C, Vitamin B12, iron).

The flash presentation will be composed of three parts: (1) principles of the algorithm used to align food ontologies using FoodOn, (ii) assessment of the algorithm using a Gold Standard specifically realized for this work, (iii) use case assessment consisting in finding in USDA food source values associated with nutrients vitamin C, vitamin B12 and iron when they are not known in Ciqua for a given food. A complete description of this work can be found in [10].

References

- [1] Pehrsson, P. and Haytowitz, D. (2016). Food composition databases. In Caballero, B., Finglas, P. M., and Toldra, F., editors, *Encyclopedia of Food and Health*, pages 16-21. Academic Press, Oxford.
- [2] Ireland, J. and Moller, A. (2016). Food classification and description. In Caballero, B., Finglas, P. M., and Toldra, F., editors, *Encyclopedia of Food and Health*, pages 1-6. Academic Press, Oxford.
- [3] Dooley, D. M., Griffiths, E. J., Gosal, G., Buttigieg, P. L., Hoehndorf, R., Lange, M., Schriml, L. M., Brinkman, F. S. L., and Hsiao, W. W. L. (2018). Foodon: a harmonized food ontology to increase global food traceability, quality control and data integration. In *npj Science of Food*.
- [4] Shvaiko, P. and Euzenat, J. (2013). Ontology matching: State of the art and future challenges. *IEEE Transactions on Knowledge and Data Engineering*, 25:158-176.
- [5] LanguaL indexed Datasets (2020) The LanguaL indexed Datasets. http://languaL.org/languaL_indexed_datasets.asp
- [6] INFOODS FAO (2020). International Network of Food Data Systems. <http://www.fao.org/infoods/infoods/fr/>
- [7] EUROFIR (2020). European Food Information Resource. <http://www.eurofir.org/>
- [8] CiquaL (2020). French food composition table. <https://ciquaL.anses.fr/>
- [9] USDA (2020). National Nutrient Database for Standard Reference. <https://ndb.nal.usda.gov/ndb/>
- [10] Buche, P., Cufi, J., Dervaux, S., Dibie, J., Ibanescu, L., Oudot, A., & Weber, M. (2021). How to Manage Incompleteness of Nutritional Food Sources?: A Solution Using FoodOn as Pivot Ontology. *International Journal of Agricultural and Environmental Information Systems (IJAEIS)*, 12(4), 1-26. <http://doi.org/10.4018/IJAEIS.20211001.0a4>