Enhancing childhood nutrition data with open biomedical ontologies

Jonathan P. Bona [1], Alexander D. Diehl [2], Alexander P. Cox [2], Aaron S. Kemp [1] and Linda Larson-Prior [1]

[1] *Department of Biomedical Informatics, University of Arkansas for Medical Sciences, Little Rock, AR, USA*
[2] *Department of Biomedical Informatics, University at Buffalo The State University of New York, Buffalo, NY, USA*

Keywords
Nutrition, neuroinformatics, biomedical ontologies, neuropsychological testing

Abstract

This poster reports our ongoing work using open biomedical ontologies to organize and analyze a diverse longitudinal dataset with elements pertaining to childhood nutrition and development. This work is part of a larger research project that seeks to advance understanding of dietary influences on psychological and neuropsychophysiological development and function in children.

Our set of 124 data elements for 600 patients includes: infant diet groups (breast milk, soy-based formula, milk-based formula), sex, size measurements (weight, length, head) at birth and at regular intervals, maternal and paternal IQ, parental socioeconomic status at yearly intervals, and a variety of mental, language, and motor developmental scores at regular intervals.

Our analysis of these data aims to identify discrete phenotypic cohorts and predict cognitive outcome measures over time. Exploratory machine learning analyses of the raw data highlighted the need for additional feature engineering to extract meaningful information. Toward this end we have deployed biomedical ontologies and semantic representations that capture connections among these data.

We construct and use explicit representations of background knowledge from relevant domain ontologies and have developed an application ontology with a small number of unique terms. These include *diet specification*, a "plan specification for a diet plan that is realized by a planned process that has as its part feeding behavior involving one or more specific foods," and *diet group role* which, like *OBI:study group role*, inheres in a population and is realized by implementation of a planned process following a specification.

Developmental measures are of particular relevance to our project aims. The Arkansas neuroinformatics group collaborates with the Neuropsychological Testing Ontology (NPT) group on related projects. Based on our preliminary modeling, we will formally define and request terms to be added to NPT for an array of childhood tests in our data, including: Preschool Language Scales (PLS) receptive and expressive language scores; Bayley Scales Of Infant and Toddler Development (BSID) mental and motor developmental scores; Reynolds Intellectual Assessment Scales (RIAS) verbal, non-verbal, and composite scores; and Symptom Assessment-45 Questionnaire: Depression.

We have converted our raw data into ontology-aligned representations in RDF/OWL using custom-built Python programs, using the OBO-ROBOT tool to extract modules of terms from other ontologies. We store these enhanced data in a triple store for reasoning and query.

In addition to providing its own inference mechanisms for reasoning about the data and exposing new connections, the semantic enhancement of these data will be used in feature engineering and selection in support of further machine learning analyses.

This project is supported in part by the Arkansas Children's Nutrition Center under USDA-ARS 6026-51000-010-06S.

International Conference on Biomedical Ontologies, September 25-28, 2022, Ann Arbor, Michigan
EMAIL: jpbona@uams.edu