CHEAR

Title - Encoding of Epidemiological Studies in Ontologies

Abstract - Epidemiology is the study and analysis of the patterns, causes, and effects of health and disease conditions in defined populations[]. Surveys[] (or Data Collection Instruments) are one of the most popular ways of collecting data for an epidemiological study. Most of the times there is a need for comparing two or more studies to understand the motivation behind them, the methods used to conduct the studies, and the content of the surveys used in these studies.

Structuring the data becomes essential also when we want to use information from current or past studies, to help make decisions in the current project or about potential new studies in the future. For these reasons we need a way to encode surveys in a formal structure like an ontology. In this paper, we define an ontology for the representation of epidemiological study questionnaires. We also present a novel way to encode questions in a survey, so that information from past surveys can be leveraged and reused in any new studies that may be conducted. This would greatly help in expediting the process of creating surveys and the epidemiologists can focus on evaluating the results from the survey to draw conclusions on the study being conducted.

Introduction -

Studies and Surveys are one of the most popular ways of collecting data when conducting an epidemiological study project. The answers/knowledge collected from these instruments/platforms are invaluable resources for Epidemiology study and analysis. Every year there are millions of forms filled out by patients. Health professions and researchers, but many of these data remain unused due to the lack of structured data/metadata collection methodology. In this paper, we study the task of Study and Survey encoding in Epidemiology. Our goal is to normalize all the questions encoding and provide a standard approach for Studies and Surveys knowledge encoding.

Use Case -

The purpose of this use case is to develop an ontology to compare and contrast two studies with one goal of determining how study results might meaningfully be combined. One study examined relationships between different body size measurements and asthma in ethnic minority children[2] while the other examined the association between prenatal and postnatal maternal stress and wheeze in 417 children enrolled in a prospective birth cohort in Mexico City[3].

During these two studies, a collection of variables were created and used to address hypothesis proposed by epidemiologists. For instance, smoking in the home increases the risk of asthma in the study participant; or, increased body size is associated with increased risk of asthma in the study participant. The data for these variables were collected by means of survey and anthropology measurement. We encoded these variables with their related concepts into our ontology, such as body composition calculations, body measure, disease, symptom, instrument, question and response. After that, we answered a set of competency questions using our ontology.

Technical Approach -

Conceptual model link: <http://owlgred.lumii.lv/online_visualization/dm62>

Ontology file link: <https://goo.gl/XnZBaV> (base ontology)

<https://goo.gl/ZLGaNI> (individual ontology)

The 2 major contributions of our ontology are :

* Body Measure and Body Composition Calculations information based on the instruments used.
* Encoding questionnaires used in different studies, and the questions asked in these questionnaires.

Methodology :

In Figure 1, according to our definition, a Body measurement is an entity that has been measured using an Instrument. A Body Composition Calculation is a class defined in our ontology, as an entity being calculated from a body measure or another body composition calculations. So we used an object property called “calculatedFrom” to map body measure to body mass calculations. We also use the “measures” object property to map an instrument to a body measure. Both of these object properties help in reasoning, when information is needed in the above mentioned domain.

Encoding questions was a major factor in our ontology. We encoded the whole question as seen in a survey, as an instance. We have a question class, with different subclasses based on the kinds of questions asked. We also have a subject class, which has many subclasses based on the topics that the questions address. The subject class is highly granular, and we use the object property “hasSubject” to map each question (instance) to an instance of type subject. This helps the user gain information on the different questions asked in a specific survey for a specific study. This property is very useful in comparing topics and questions used in different studies. This helps gain information about many different studies and the approaches they use.

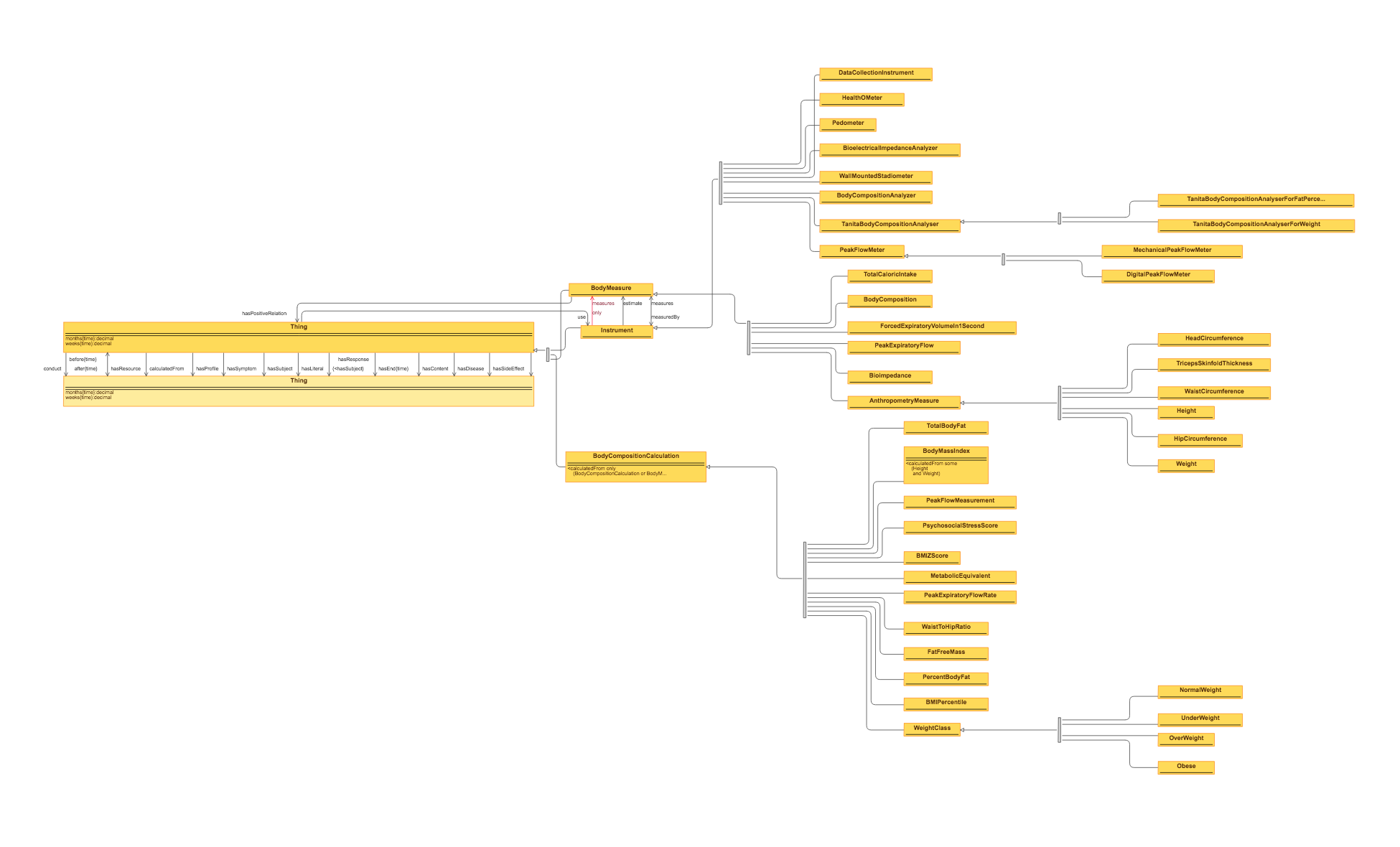


Figure 1

Discussion -

Value of Semantics -

Semantics can be used to help ensure consistency and support disambiguation in the use of terminology. In our use case, the two studies we encoded contain terms that are related to childhood asthma and obesity. Though having different mentions in the two studies, some of the terms actually refer to the same things. By the use of the ontology, the semantics of these terms can be standardized and thus properly understood by human beings and machines. Then we can know what variables are currently shared by the two studies and what variables in one study can be somehow reused in the other one. This will potentially promote the integration of data or results from different studies and help epidemiologists to discover new relations between factors. Another benefit brought by semantics in our use case is that the ontology helps people who are not domain experts to better understand the studies.

Project website - <https://tw.rpi.edu/web/Courses/Ontologies/2016/projects/chear>

Support your claims -

We define an ontology for the representation of epidemiological study questionnaires. We also present a novel way to encode questions in a survey, so that information from past surveys can be leveraged and reused in any new studies that may be conducted, i.e. new question/topic/category can be easily added as new instance with a reference to an existing concept (or a newly created concept). To evaluate our ontology, we use it to properly answer a set of competency questions using sparql queries and validate the syntax and corresponding logical semantics of the ontology for correctness using syntax and consistency checking tools.

Future Work -

In the future, we plan to collectively expand the ontology and make it applicable to general domain and introduce it to general audience. In addition, we expect to integrate our work into a larger ontology we designed for the CHEAR (Child Health Exposure Analysis Repository) project.

Works Cited

[1]"Epidemiological Survey." *TheFreeDictionary.com*. Web. 06 May 2016.

[2]"Epidemiology." *Wikipedia*. Wikimedia Foundation. Web. 29 Apr. 2016.

[3]"Questionnaire." *Wikipedia*. Wikimedia Foundation. Web. 06 May 2016.

[4]Rosa, Maria José, Allan C. Just, Marcela Tamayo Y Ortiz, Lourdes Schnaas, Katherine Svensson, Robert O. Wright, Martha María Téllez Rojo, and Rosalind J. Wright. "Prenatal and Postnatal Stress and Wheeze in Mexican Children." *Annals of Allergy, Asthma & Immunology* 116.4 (2016). Print.

Vangeepuram, Nita, Susan L. Teitelbaum, Maida P. Galvez, Barbara Brenner, John Doucette, and Mary S. Wolff. "Measures of Obesity Associated with Asthma Diagnosis in Ethnic Minority Children." *Journal of Obesity* 2011 (2011): 1-9. Print.

"Epidemiologic Study." *TheFreeDictionary.com*. Web. 06 May 2016.