A data structure for the implementation of referent tracking systems

Lauren Wishnie [1] and Alexander Diehl [1]
[1] *University at Buffalo, 77 Goodell Street, Buffalo, NY, USA*

Abstract
Our goal is to implement a Referent Tracking paradigm in tandem with a fully axiomatized Alzheimer's Disease Neuroimaging Initiative (ADNI) knowledge base.

Keywords
Referent tracking, Alzheimer's disease, ontology, knowledge base

Referent Tracking (RT) seeks to address a design flaw often seen in databases wherein the majority of assertions lack an explicit reference to what the assertions are about. *Referents* are the entities that assertions are about. RT originated as a response to this problem in the context of Electronic Health Records (EHRs) but the problem exists in research-derived data as well. One of the primary tenets of the RT paradigm is the assignment of unique identifiers (here referred to as RUIs) to all referents and the assertions made about them, so that changes in the data mirror either changes in reality itself, resolves an error, or comes from changes in the users' understanding of reality. In RT, these changes are tracked using predicates called *tuples*. There are a total of eight tuple types, each of which holds a specific type of information related to the assertion. Importantly, the various tuples allow us to make explicit changes in relations between particulars and particulars as well as particulars and universals. Thus, RT frameworks are designed to work in tandem with realism-based ontologies. Our goal was to create a relatively simple framework for generating tuples as part of tracking changes in the Alzheimer's Disease Neuroimaging Initiative (ADNI) data. The RT implementation will be supported by the ADNI Ontology, which is currently in development and will be annotated with the ADNI data, forming an ADNI knowledge base. This RT data structure was built in Python. Each tuple type is generated by a function which is specifically designed to adhere to that type's structure. The data structure accepts axiomatized assertions, which express relationships between different types of entities, including real-world referents and ontology terms. The program then places the assertion into the appropriate slot in the tuple template. Next, the tuples are compiled into lists, which are then converted into dataframes using the .DataFrame() function in the Pandas library. The creation of each tuple is timestamped using the datetime library. The UUID library is used to auto-generate new RUIs for every new tuple and any other relevant portions of reality. Implementations of the RT paradigm vary depending on what kind of data are in the database and how changes are currently done. In the case of this work, the modes of input for the data structure have been developed to track changes in the ADNI data. The ADNI data are longitudinal, with several measurements per participant across multiple years. A major goal of this work is to create an RT system that allows the tracking of events in the evolution of a participant's medical history. Possibilities for future work include using RT databases as a basis for the visualization of patient timelines and other data elements.

ICBO 2022, 07–01, 2022, Buffalo NY, USA
EMAIL: lmwishni@buffalo.edu (A. 1); addiehl@buffalo.edu (A. 2); ORCID: **0000-0002-7245-3450** (A. 1); **0000-0001-9990-8331** (A. 2)