

Steps Toward the Automated Assembly of Knowledge Bases from Text

**Joseph Marr^{1,5,6*}, Hinduja Nallamala², Saman Shahid³,
Daniel Spindel², Nawaf Al-Dhelaan⁴, Edward Wegman^{5,6}**

1. SYNTASA, Inc.

2. George Mason University, College of Science, Department of Biology

3. George Mason University, College of Humanities and Social Sciences, Department of Neuroscience

4. George Mason University, Volgenau School of Engineering, Department of Computer Science

5. George Mason University, School of Physics, Astronomy and Computational Sciences (SPACS)

6. George Mason University, SPACS, Center for Computational Statistics and Data Science

*Contact author: joseph.marr@gmail.com

Keywords: Text analytics, Semi-structured data, Export systems

The IMPLICATION project seeks a generalized methodology to extract instances of linked causation statements from text. Linked causation statements are statements of the form $A \rightarrow B$ and $B \rightarrow C$ (or, alternatively, $A \rightarrow B \rightarrow C$), where A is the antecedent (cause) for consequent (effect) B , and B is the antecedent for C . Here, we focus on examples of explicit causation statements that appear in biomedical text, and that are indicated by phrases we have designated “causation signals.” The syntax and semantics of several causation signals will be discussed, along with their computational implementation. The results of this study bring us a step closer toward realizing an automated system for knowledge assembly, where knowledge is here understood to be comprised of sets of linked production rules.