

# DataDeps.jl

and other foundational tools for data driven research  
(Especially NLP)

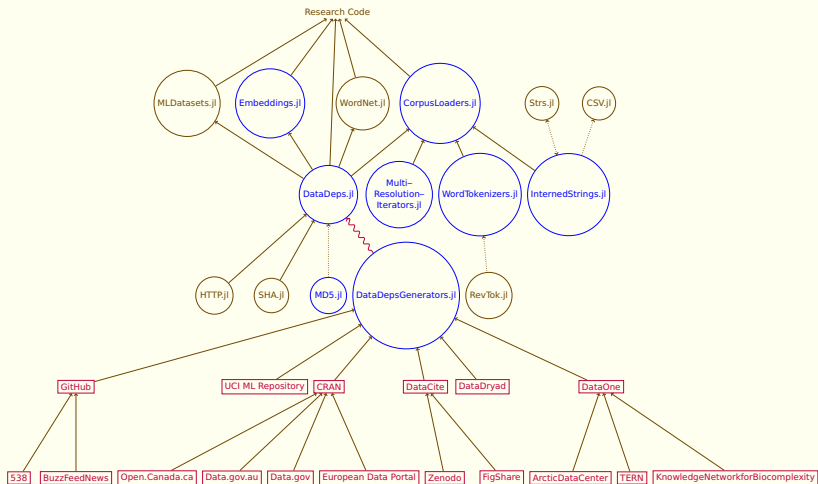


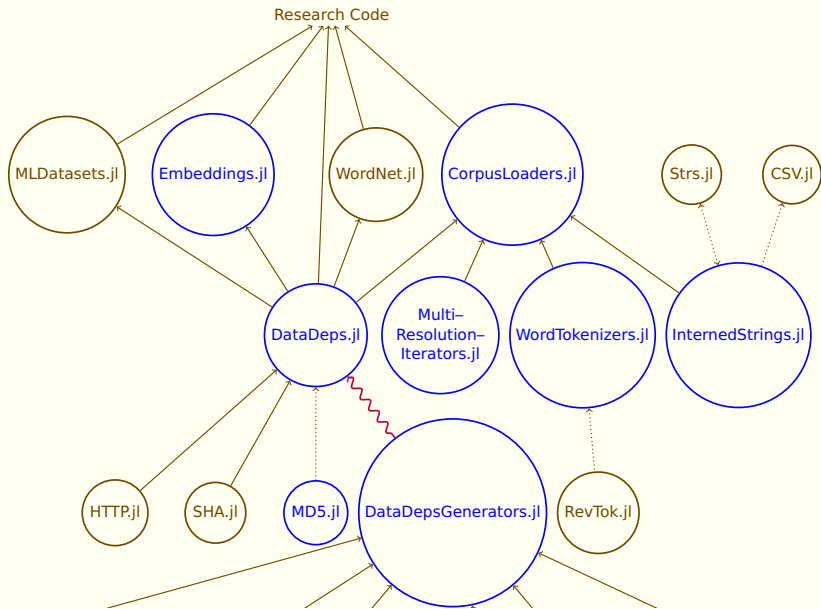
**Lyndon White**

School of Electrical, Electronic and Computer Engineering  
The University of Western Australia



THE UNIVERSITY OF  
**WESTERN  
AUSTRALIA**





# Vabdwakke's 6 Degree's of Replicability

1. The results cannot seem to be reproduced.
2. The results could be reproduced by, requiring extreme effort.
3. The results can be reproduced, requiring considerable effort.
4. The results can be easily reproduced with at most **15 minutes** of user effort, requiring some proprietary source packages (MATLAB, etc.).
5. The results can be easily reproduced with at most **15 min** of user effort, requiring only standard, freely available tools (C compiler, etc.).

Vandewalle, Kovacevic, and Vetterli (2009), "Reproducible research in signal processing"

# What happens when I try and reproduce someone's research code?

**1min** Find the website from the paper, and download the code

**2min** Read enough of the README to get rough bearings

**1min** Find out where to get the data from and download the data

**2min** Try and remember how to use `tar -xzfvalphabetsoup` etc.

**2min** Workout how to tell script where data is

**2min** Setup any software dependencies etc.

**3min** Run the code and make sure it isn't crashing etc.

**2min** Interpret the output