It started as a toy problem for flexing my new skills with the R programming language: how might the nuances of philosophical texts be represented visually, and efficiently, for students?

My resulting tool will be one of the ways in which I am making the abstract and complex accessible to students, and therefore part of the public conversation. In the meantime, my focus has been on creating new ways to mine text using the quanteda package in R.

The basic unit of analysis is an ideme: my term for a meaningful phrase or word that cannot be used to understand the text without context, like "an evil." An ideme alone cannot tell us what the text says, much less the sense in which a word with multiple meanings, like "evil," is used. The sentence that contains the ideme "an evil" could mean something bad happening to someone, or an evil person.

With these issues in mind, simply filtering out "important" uncommon words like "evil" tells us very little about the nuances of an argument within a text. But how to find these idemes?

From reading the text of William James's *The Will to Believe*, I figured out a list of tag words, like "because" and "since." These are stored in a dictionary, to tell R where there is an ideme in a particular sentence. I borrowed the idea of tag words from another method of parsing linear information: DNA sequencing. Genetic sequencing machines break up genomes by tagging particular combinations of nucleotides, and my program does the same with the linear philosophical texts.

Stored in a separate dictionary in R are operator words that I used similarly, to tell R that the idemes on both sides are related to one another in particular ways. I borrowed the idea from computer logic operators like "and" and "or."

To ensure that the visualization displayed is comprehensible, I have to learn to put the text back together; my work so far has taken it apart and divided it into idemes. To put the text back together is a harder problem; it requires me to find an overarching structure that will make idemes comprehensible within a custom type of data visualization. I may further investigate how DNA is spliced in bioengineering technologies like CRISPR, since formal logic so far gives no indications of words that put arguments together in the systematic way required for visualization.

How might the nuances of philosophical texts be represented visually, and efficiently, for students, from text mining in R?

The basic unit of analysis is an ideme: my term for a meaningful phrase or word that cannot be used to understand the text without context, like "an evil." An ideme alone cannot tell us what the text says, much less the sense in which a word with multiple meanings, like "evil," is used. The sentence that contains the ideme "an evil" could mean something bad happening to someone, or an evil person.

Because a basic unit of analysis is not enough to create a cohesive picture of any phenomenon, the next stage is to systematically link the idemes together.