

## Webscraping & Word-to-Vectors: Making sense of democratic opinions

The idea of this project is to use Deep Neural Nets that implement Word-to-Vector analysis solutions in order to make sense of text formulated in natural language. For example, public policy reforms like in health care and taxes consist of a large variety of issues and opinions. The idea is to explore how deep neural nets can be used to analyze written text to map out the qualitative opinion landscape of the different issues involved in these issues. This will require to convert qualitative text into quantitative numbers, and Word-to-Vector approaches have been very successful here. The task is two-fold:

- 1) Chose one politically relevant topic (e.g. health reform or tax reform) and create a webscraper to obtain op-ed / opinion related newspaper articles from different newspapers: New York Times, Breitbart, NPR, MSNBC, etc.
- 2) Explore how word-to-vector applications can be used to summarize the opinions and to make quantitative sense of them (e.g. <https://nlp.stanford.edu/projects/glove/> or <https://www.tensorflow.org/tutorials/word2vec> ). Can we create a more comprehensive picture about the “opinion landscape” with regard to these issues?

---

Martin Hilbert (Dr; PhD) | Associate Professor

Dept. of Communication | University of California, Davis

[www.martinhilbert.net](http://www.martinhilbert.net)