

Automating the Comlaw Database: Exploring the limits and possibilities of replacing human coders

Authors: Matthew Gabel and Jeff Ziegler (Washington University in St. Louis)

Abstract:

The Comlaw database provides a common coding template for describing the context and outcomes of constitutional review across courts and time. In a pilot project, we have shown the template is reasonably successful for systematically coding relevant information about the content and context of constitutional rulings for 48 courts involving a broad array of languages for one year (2003). However, the coding template relies on human coding. This has severely limited the expansion of the database and, consequently, its value for researchers. In this paper, we evaluate the potential for populating the Comlaw database through an automated process based on text analysis of the publications of the courts. Automation faces a variety of hurdles related to the formatting of documents and the language used. As a test case, we chose the French Constitutional Council which has ruling and supporting documents available electronically for a long time-series and with some potentially challenging features related to variation in the format/type of rulings (e.g., *ex ante* vs. *ex post*). The goal of the exercise is to learn how much of the coding can be accomplished through automation. And, where automation cannot directly identify the information needed to complete the database, we hope to define an algorithm that can isolate the relevant text that a human coder would need to complete the coding.