CFR_abstract

R. Schwinn

1/30/2018

A huge amount of effort is spent analyzing new rules during the promulgation process, but little attention is paid to analyzing the existing stock of rules. The CFR Miner uses a variety of natural language processing (NLP) techniques to enable users to easily and effectively analyze the Code of Federal Regulations (CFR). Its interactive visualizations let users browse the 4.5 million word corpus with a swipe of the mouse. Force vector maps and collapsible trees reveal structure and accentuate the interconnections between rules. Summarization algorithms provide succinct k-sentence summaries of any granularity of CFR content ranging from paragraphs to volumes. Users can upload, copy & paste, or provide links to their own content, such as business plans or proposed rules, to identify related CFR entries. Classification and clustering analysis techniques, also known as supervised and unsupervised learning, respectively, identify explicit and implicit CFR features by identifying the latent structure of the corpus. The CFR miner is primarily written in R and XPath but relies heavily upon C++, javascript, HTML, and CSS through Shiny and various R packages.