A Right to Access Implies A Right to Know: An Open Online Platform for Research on the Readability of Law

Michael Curtotti, Eric McCreath

Abstract

The widespread availability of legal materials online has opened the law to a new and greatly expanded readership. These new readers need the law to be readable by them when they encounter it. However, the available empirical research supports a conclusion that legislation is difficult to read if not incomprehensible to most citizens. We review approaches that have been used to measure the readability of text including readability metrics, cloze testing and application of machine learning. We report the creation and testing of an open online platform for readability research. This platform is made available to researchers interested in undertaking research on the readability of legal materials. To demonstrate the capabilities of the platform, we report its initial application to a corpus of legislation. Linguistic characteristics are extracted using the platform and then used as input features for machine learning using the Weka package. Wide divergences are found between sentences in a corpus of legislation and those in a corpus of graded reading material or in the Brown corpus (a balanced corpus of English written genres). Readability metrics are found to be of little value in classifying sentences by grade reading level (noting that such metrics were not designed to be used with isolated sentences).

Keywords

readability; legislation; legal informatics; corpus linguistics; machine learning

References
