

How Technology Can Help Reducing the Legal Burden

Tom van Engers^a, Ron van Gog^b and Arian Jacobs^c

^a*Leibniz Center for Law, University of Amsterdam,*

^b*Belastingdienst, The Netherlands,*

^c*Rulewise, The Netherlands.*

Abstract. This short paper explains how relative simple technology can help governments to reduce the legal burden.

1. Introduction

The Dutch Tax and Customs Administration (DTCA) as any other public administration publishes decisions in how to interpret the discretionary room that always is present in legislation in the form of policy directives which have a legal status. As part of a national programme aiming at reducing the administrative burden and reducing (perceived) regulatory pressure (see also [1]), the DTCA started a task force aiming at reducing the number of these policy directives from about 1400 policy directives to about half. We supported this project with technical solutions that help the project manager and her crew to intelligently prune the set of policy directives and consequently meet her project's goals. The technology used wasn't exactly rocket science, consequently we don't claim that we present highly innovative AI-based techniques, but we want to show a useful application of AI-based techniques in a practical legal setting.

2. The Clustering Process

Policy directives are a solution to the unintended openness of legal concepts and serve both in diminishing discretionary room for civil servants and improving transparency and equality before the law for the citizen. Through the years many of such policy directives have been published. It is both in the interest of the citizens (taxpayers) as that of the civil servant themselves that the number of such policy directives doesn't become too large.

For clustering we first created a vector representation of the policy directives. Every dimension is associated with a word (normally from a list of keywords) in the source document. The number of dimensions is thus determined by the number of words n . Every coordinate is determined by the weight of the corresponding words (dimensions). The weight of a word w in document I is calculated as:

$$weight(w,i) = f_{wi} \log\left(\frac{N}{n_w}\right)$$

In this formula f_{wi} is the frequency in which the word w occurs in document i , N the number of documents in the corpus and n_w the number of documents that contain the word w . This way the weight of a word (in a single document) is proportional to the number of occurrences in the document. The factor $\log(N/n_w)$ causes that the weight is weakened proportionally to the occurrences in other documents in the corpus. A word occurring in all documents consequently will have a weight 0 in all documents (no significance). We normalised the vectors in order to make them comparable. Meaningless words such as determiners or propositions, are included as dimension but will be factored out by $\log(N/n_w)$.

The correspondence between documents can be expressed as the dot product of the documents' vectors. The dot product of two normalised vectors is equal to the cosine of the angle between those vectors. A value 1 means that the angle between the vectors is equal to 0 degrees which implies that the documents have (almost) the same meaning.

We clustered the policy directives (using 0,7 as cut-off value) which resulted in 380 clusters with an average size of 2.74 documents and a maximum cluster size of 58. We then presented the clusters to the experts. These experts were surprised that we were able to offer them a set of correlated documents without having any legal knowledge about the complex legal topic that was described within them. We even included documents that they overlooked, but were relevant.

3. Conclusions

The project mentioned in this short paper is just an example of an attempt to reduce the legal burden. The approach explained in this paper proves that in order to support legal experts it is not always required to use rocket science. Using existing technology sometimes with little extensions or used in a slightly different way can already provide a lot of help in difficult legal tasks. The full paper can be found on www.LeibnizCenter.org.

References

- [1] Arendsen, R. and Van Engers, T.M., An Empirical Study on Law Enforcement and Data Reporting Strategies, in: Proceedings of E-Government Workshop in conjunction with JURIX 2003, Palmirani, M. et al., Universitatsverlag Rudolf Trauner, December 2003.
- [2] Van Engers and T.M., Legal engineering: A knowledge engineering approach to improving legal quality. In J. Padget, R. Neira, and J.L. De León, editors, *eGovernment and eDemocracy: Progress and Challenges*, pages 189-206. Instituto Politécnico Nacional Centro de Investigación en Computación, 2004. ISBN 970-36 0152 9.
- [3] Van Engers, T.M. and Boekenoogen, M.R., Improving legal quality - a knowledge engineering approach. In Galindo, editor, *International Review of Law Computers & Technology*, volume 18. Carfax Publishing, March 2004. ISSN 1360-0869.