

Analyzing Legislative Burden Upon Businesses Using NLP and ML

Summary

In this workshop, we'll first describe the legislative/business context for the initiative, then walk attendees through the technical implementation. The work will be conducted by combining various techniques from the NLP toolbox, such as entity recognition, part-of-speech tagging, automatic summarization, and topic modeling. Work will be conducted in Python, making use of libraries for NLP such as spacy and nltk, and the ML library scikit-learn. We will also showcase interactive dashboards which have been created using the BI tool Qlik to allow exploration of the results of the analysis.

Requirements

Github repository

All required files can be downloaded from [this](#) github repository.

A complete list of the libraries used in the workshop can be found in `requirements.txt`. To install using pip run

```
pip install -r requirements.txt
```

NLP libraries

If you've never used `nltk` or `spacy` you might need to download a few extra resources.

Spacy english model

In a terminal window run

```
python -m spacy download en
```

More details on spacy's models can be found [here](#).

nlk corpora

In the python shell run

```
import nltk
nltk.download('stopwords')
nltk.download('wordnet')
```