**Bill Classifier/Prediction (BCP) Tool Design Document**

# Project Summary:

Currently ‘Council Data Project’ website hosts video recordings of City of Seattle council meetings and their transcripts. BCP tool takes ‘Bill title’ as input and predict if a new bill will pass or not based on the data mining from meeting transcripts. Training data will include the sentiment analysis on the transcript text field and if the bill actually passed in the council.

# High level BCP tool design:

This tool will implement Natural Language Processing algorithm to predict if a new bill in council will pass or not. The transcript text from the CDP website will be saved in the backend, and BCP tool will pick data directly from backend. Before implementing NLP algorithm, we will use sentiment analysis in following steps:

1. Feature vectors: Unique set of words
2. Vocabulary size: Number of unique words
3. Convert to data vector: Number of occurrences of each word
4. Interpret weights using Logistic regression

## High level BCP Tool Data flow diagram:

Database saves transcript text

Transcript on Council Data Project Website

BCP Tool

# Backend design:

* Dedicated table to save transcript data and related transformed data
* There are no full stops, commas etc in source text/transcripts
* Named Entities can be saved in separate columns so that search is fast
* Day and date of the meeting/transcript
* Classify sentences in the transcript as in favor or against the bill and comparing if the bill actually passed or not.
* Sentiment word dictionary?

# Scope:

Prediction of result will be based on the transcript texts on CDP website collected from insert date here only.

# Out of scope:

# Schedule:

# Assumptions: