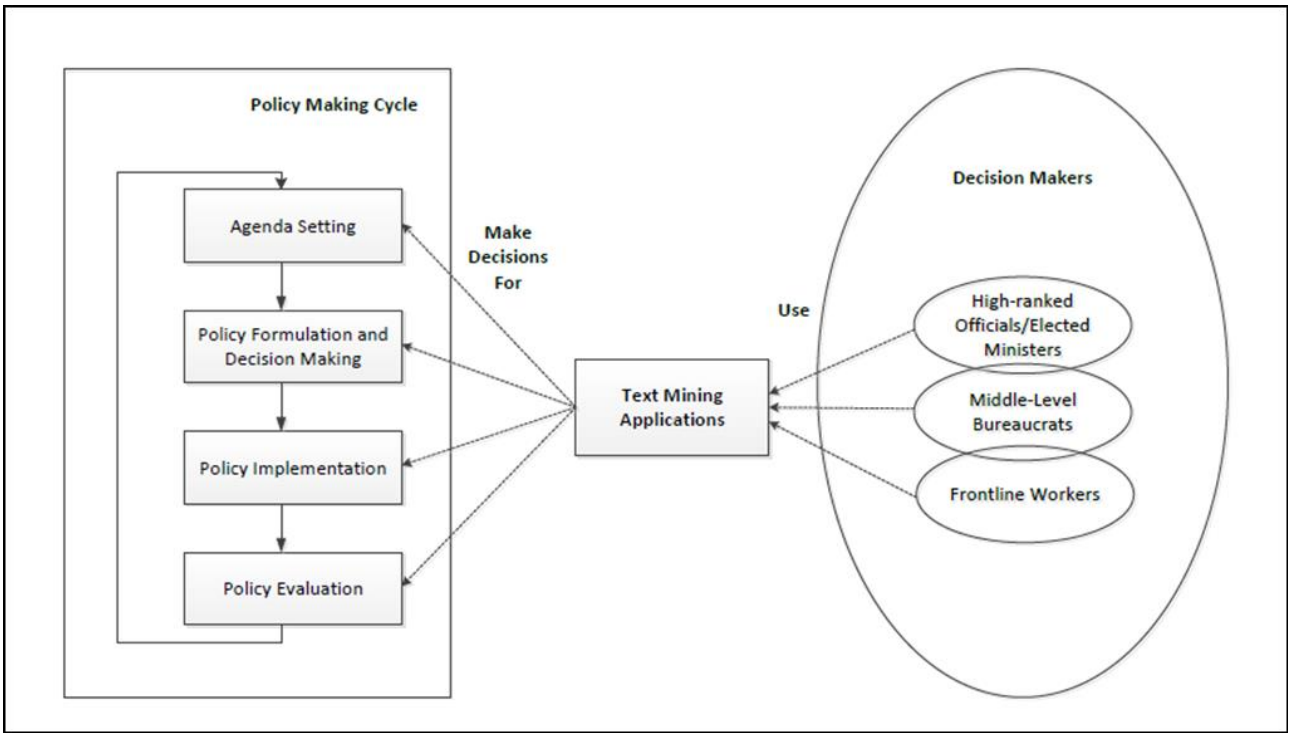
# Final Project Prospect of text mining

Shinya Takatani / UCID: 12147306

Text Mining or Natural Language Processing have huge potential to be utilized in public policy. According to the review paper by E.W.T.Ngai and P.T.Y.Lee in 2016, they developed a conceptual framework which has four steps in policy making. It includes the policy making cycle similar to PDCA, which consists of 1. Agenda Setting 2. Policy Formulation and Decision Making 3. Policy Implementation and 4. Policy Evaluation. The text mining technique will enable public sectors to enhance the quality of the cycle and reduce the cost by saving time for the process.



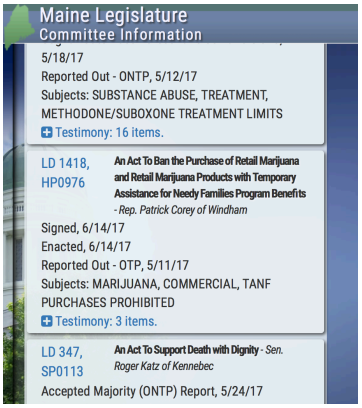

In my project, I would like to propose a total package to improve each policy making process with text mining skills which we will learn in this class. I will apply the technique for the example of "Temporary Assistance for Needy Families (TANF)" in the following steps. The program is designed by the federal government (Department of Health & Human Services, Administration for Children & Families) to help needy families to work and care children self-sufficiently. And the grant is given to state governments so that the program is operated by themselves.

## 1. Agenda Setting

At the first step, policy makers are interested in gauging the attention and sentiment of stakeholders including the public. I can know the trend or the forecast from news articles or SNS. The data could be easily scraped by the related words (for example, "TANF"), using Twitter API and NY Times API. Fortunately, some sentiment analysis tools are available for Python tough some of them do not seem free of charge. I could speed up my analysis by leveraging one of them.

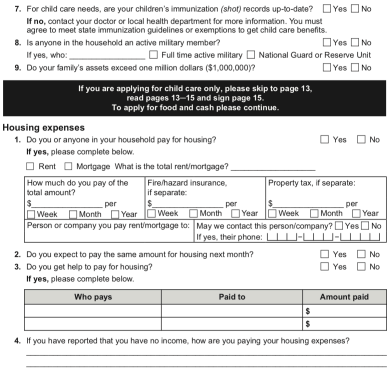
2. Policy Formulation and Decision Making

The general rules or regulations of the law have been formulated by the federal legislature, but the details of the program depend on state legislatures and the programs are operated by the state governments. We can find each program site of states from here, but they are so messy and diversified that I cannot immediately reach to correct information which I want. Therefore, I should see the decision making process on the website of each state legislature since we already dived into scraping documents on some state legislatures. For example, Maine legislature's committee on Health and Human Services has some agendas on TANF with testimony pdf files. My goal in this step is to download files for agendas in each state and classify the opinions or the amendments, distinguishing opponents from proponents by assessing their statements. If I can, I would like automatically provide a brief summary including reasons for the position of each statement.




3. Policy Implementation

The automatic screening or the error detection for application documents is one of prospective public applications of text mining. Many frontline workers are involved in this kind of process, so we can cut the project cost if the process is automated. Unfortunately, I cannot have access to the real application document database of TANF because they include much personal information. Here, I will create a simple mock-up which detects handwriting errors in the documents, and I will also check the function by a sample which I fill in by myself, using an official format. I will take advantage of one OCR application such as Google Cloud Vision API.


7. For child care needs, are your children's immunization (*shot*) records up-to-date? ☐ Yes ☐ No
**If no,** contact your doctor or local health department for more information. You must agree to meet state immunization guidelines or exemptions to get child care benefits.
8. Is anyone in the household an active military member? ☐ Yes ☐ No
If yes, who: _____ ☐ Full time active military ☐ National Guard or Reserve Unit
9. Do your family's assets exceed one million dollars ($1,000,000)? ☐ Yes ☐ No

**If you are applying for child care only, please skip to page 13, read pages 13–15 and sign page 15. To apply for food and cash please continue.**

**Housing expenses**

1. Do you or anyone in your household pay for housing? ☐ Yes ☐ No
**If yes,** please complete below.
☐ Rent ☐ Mortgage What is the total rent/mortgage? _____

| How much do you pay of the total amount? | Fire/hazard insurance, if separate: | Property tax, if separate: |
|---|---|---|
| $ _____ per ☐ Week ☐ Month ☐ Year | $ _____ per ☐ Week ☐ Month ☐ Year | $ _____ per ☐ Week ☐ Month ☐ Year |
| Person or company you pay rent/mortgage to: | | May we contact this person/company? ☐ Yes ☐ No If yes, their phone: (_ _ _) _ _ _-_ _ _ _ |

2. Do you expect to pay the same amount for housing next month? ☐ Yes ☐ No
3. Do you get help to pay for housing? ☐ Yes ☐ No
**If yes,** please complete below.

| Who pays | Paid to | Amount paid |
|---|---|---|
| | | $ |
| | | $ |

4. If you have reported that you have no income, how are you paying your housing expenses?

4. Policy Evaluation

Strictly speaking, I think that this process is out of scope of this class because the program evaluation should usually be based on quantitative analysis rather than on qualitative reviews. (We learnt some methods to estimate treatment effects of programs in other courses.) However, it is valuable to pay attention to feedback comments of participants or frontline workers in existing programs. They might give some hints to decide performance measures of TANF which are mentioned in the past work by H. Hahn and P. Loprest.

As mentioned above, I can utilize text mining to make TANF more efficient. However, some processes of the package might be more challenging than I expect, so I am not sure that I could present perfect results for all the tasks in the final work. Therefore, I would focus on the second step if I don't have much time. I would like to provide some important findings on what I can and cannot do with state-of-the-art text mining rather than take efforts to write well-developed codes. I believe that this try would give us practical insights on text mining or NLP.