**Resources/Links**

* <https://www.nlc.org/sites/default/files/2019-05/StateOfTheCities%202019.pdf>
* <https://towardsdatascience.com/a-practitioners-guide-to-natural-language-processing-part-i-processing-understanding-text-9f4abfd13e72>
* <https://towardsdatascience.com/your-guide-to-natural-language-processing-nlp-48ea2511f6e1>

**Key Notes from NLC’s State of the Cities Report**

* **NLC** - the voice of American Cities, towns and villages. Works to strengthen local leadership, influence federal advocacy and drive innovative solutions
* **NLC Center for City Solutions** - provides research and analysis on key topics and trends important to cities, creative solutions, inspiration and ideas, opportunities for city leaders to connect with peers
* **State of the Cities Report 2019** – in depth analysis of 153 mayor’s state of the city (similar to presidents State of the Union) speeches
* **Methodology for 2019 Analysis:**
  + 153 speeches given from January through April 2019
  + Speeches are coded as having significantly covered a major topic if the word count for that topic constitutes at least 10 percent of the speech
  + Only code words if the mayor indicated a specific plan, goal, or impact for 2019/0r future
  + 185 subtopics that are then matched to 10 major topics
  + Mayors discussed 1-55 topics per speech with average of 21 subtopics per speech
* **Results from Analysis:**
  + Show top 10 major topics and compare to previous years
  + Focus on top subtopics within 10 major topics
  + Subtopics broken down by region
  + Subtopics broken down by population
* **Other Insights:**
  + Economic development - most-covered policy issue. 2019 74 mayors dedicated a significant portion, 2018 58 percent.
  + Health and Human Services - climbed 3 spots from previous year. 2019(46 percent), 2018 (34 percent) - increased focus on expanding parks and recreation facilities
  + Energy and environment - 2019 (41 percent), 2018 (25 percent)
  + Housing - 2019 (38 percent), 2018 (39 percent.
  + Health and Human services jumped three levels 0 mostly because of increased focus on parks and recreations. Other sub topics played and important role. References to mental health

**Potential Strategies/Notes from Staff meeting w/Kunath**

* Importing data from Nvivo to Python
* Machine Learning would be in classifying parts of speech with pre-existing 10 topics
* Cross document structural analysis. For example, do different people talk about the same issues in different orders?
* Nltk.tokenize.texttiling – text segmentation, partitioning texts into coherent multi paragraph discourse. After doing textiling you can compare the text identified among other mayors speeches to identify overlap
* Identifying future tense – look for modal verbs like will, we will. Maybe spacy has something for this. You can also collect the most frequent verbs as well as verb phrases. This way you can identify things that mayors are thinking about. Can do separation to find pharased like we did vs. we aspire to. Also look into FrameNet from Berkley, lets you figure out types of actions that verbs can suggest
* Look into techniques like info map. Maybe only in C
* Unsupervised learning may not be the best idea because we they already have subtopics. Maybe start with Naïve Bayes classification
* One issue is variation in text – they say the same thing but mean different things. Try lemmatizing the text
* Neural classifiers out of the deep learning space. The features are word vectors that you would get from spacy
* One thing to keep in mind is lexical drift – across years, using different phrasing but has the same meaning
* Bag of Words
* Topic Modeling