

Bloomberg Project

Atufa Shireen

→ Approach

- The [Federal Register API](#) is used to extract document numbers, excerpts, titles, url to articles and publication dates.
- The with a manual check on few documents, title and excerpts were not found to perfect explanation to the document. [Example](#).
- Therefore, the full document including title, introduction, summary, abstract and body was used.
- The final document is created by scraping paragraphs from each of the article using beautifulsoup, and adding the title, summary to it and removing few common articles relating to publishing guidelines.
- Text preprocessing is performed on the document,
 - Lowering the text.
 - Removing punctuations.
 - Removing spaces, tags.
 - Removing stop words and short words.
 - Removing numbers and alphanumerics.
 - Lemmatization of words
- Bag Of Words is used to create dictionary.
- The TFIDF, or term frequency–inverse document frequency, is used to extract words with weight.
- Gensim is used for creating preprocessing pipeline and topic modelling using LDA.

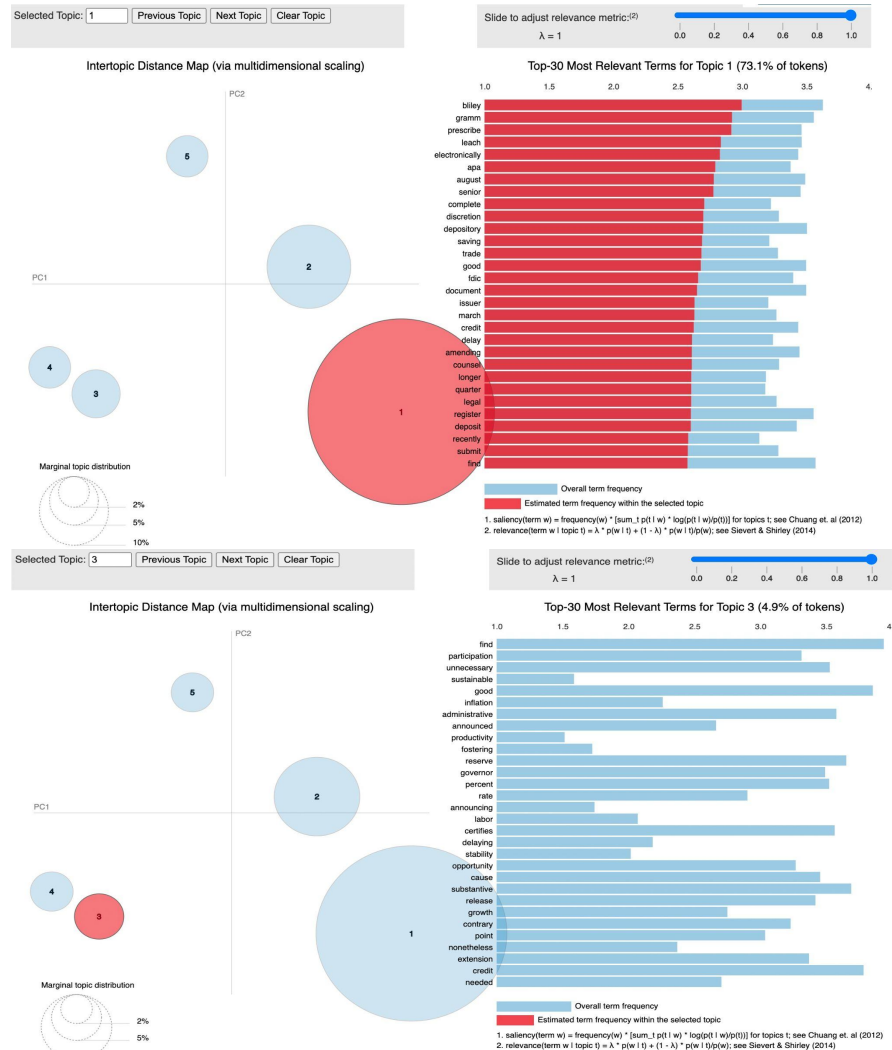
➔ What were the common regulatory priorities of these agencies from 2001 through 2006?

- **Topic:** Gramm-Leach-Bliley Act

The Gramm-Leach-Bliley Act requires financial institutions – companies that offer consumers financial products or services like loans, financial or investment advice, or insurance – to explain their information-sharing practices to their customers and to safeguard sensitive data.

- **Topic:** Sustainability

Sustainability is a societal goal with three dimensions: the environmental, economic and social dimension. This concept can be used to guide decisions at the global, national and at the individual consumer level.



→ What new topics and issues emerged in the Obama Admin (2009-2017).

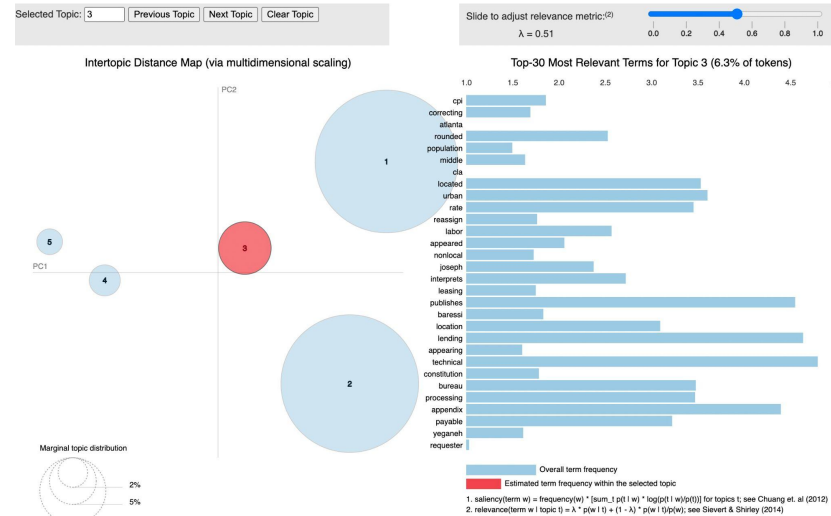
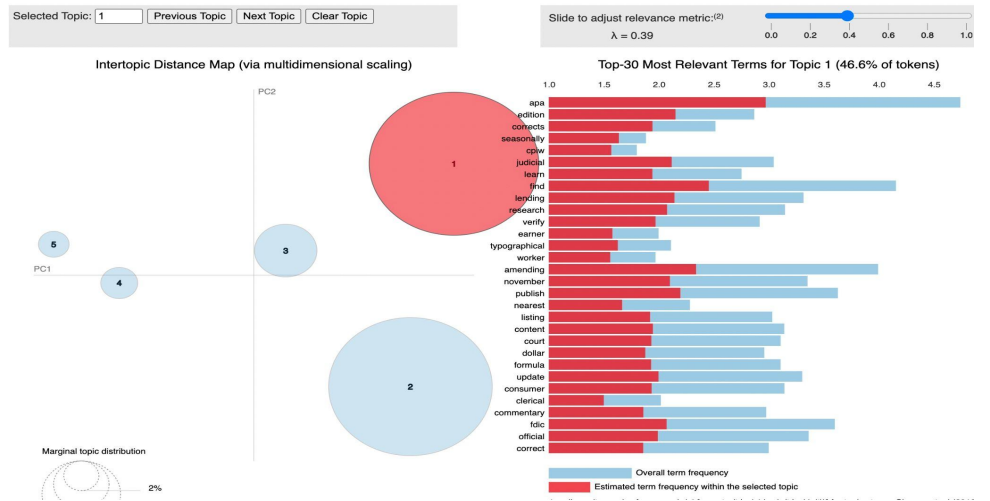
• **Topic: APA**

The topic talks about corrections or creating editions to APA or Advanced Pricing Agreements which is an ahead-of-time agreement between a taxpayer and a tax authority on an appropriate transfer pricing methodology for a set of transactions at issue over a fixed period of time.

• **Topic: CPI and CLA**

The consumer price index is the most widely used measure of inflation, closely followed by policymakers, financial markets, businesses, and consumers.

A consumer lease Agreement states, a lease contract where someone (a lessee) is leasing goods for personal use and either has an option to purchase the leased goods, or the term of the lease is over 1 year.



- Impact of the Coronavirus Pandemic (2020~2021)

- Topic: CMP**

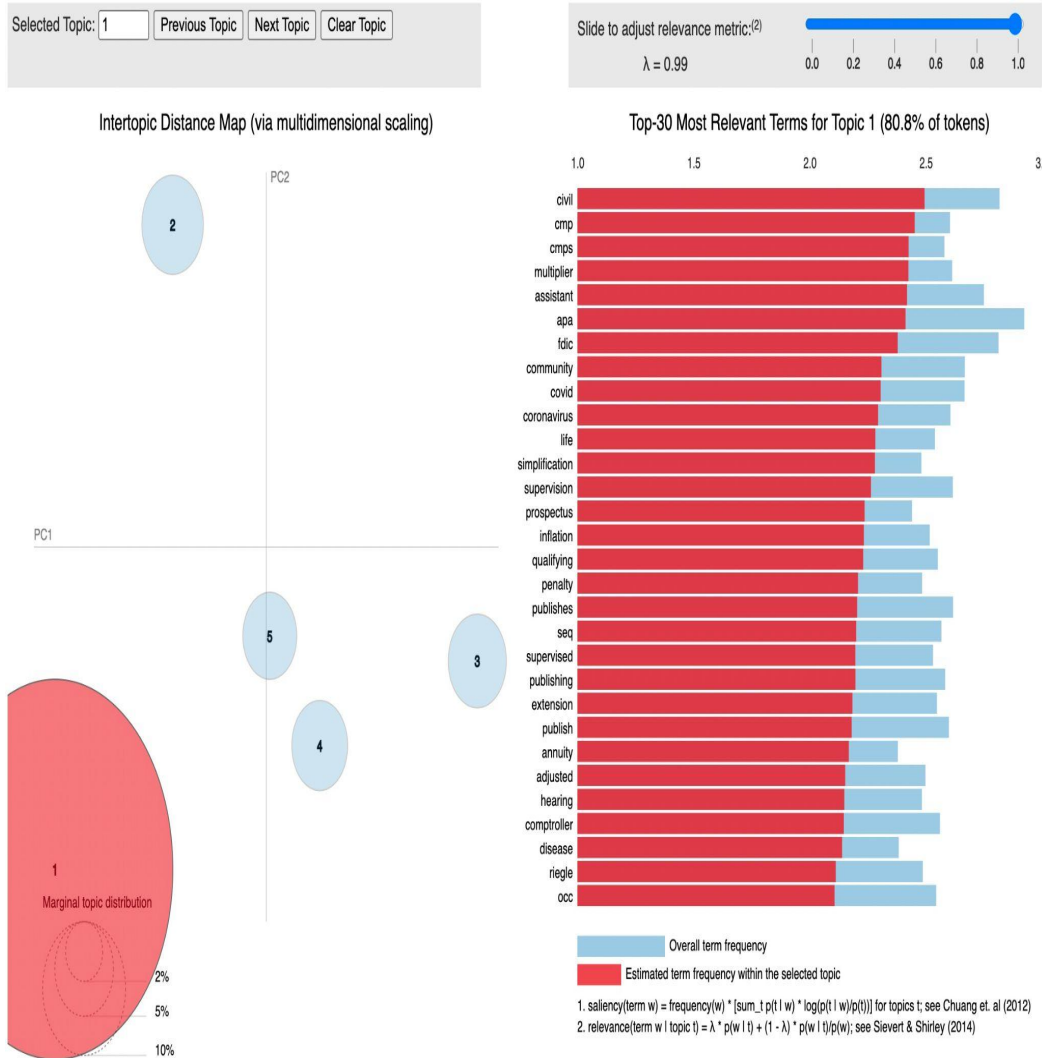
The Centers for Medicare & Medicaid Services, is a federal agency within the United States Department of Health and Human Services that administers the Medicare program.

- Topic: Coronavirus, disease**

We can also observe articles addressing the issue of coronavirus as disease and assistants that can be provided to people.

- Topic: OCC, Comptroller**

The active participation of OCC agency can be observed during this period.



→ Improvements

- Although with manual check few common paragraphs such as detailing guidelines were removed, further reduction of more common paragraphs can improve the result of result.
- Taking few or one year (~600 docs for covid) at a time shows more accurate topic details than taking 9 topics at a time(~ 1600 docs for obama admin).
- For convenience, paragraphs with >300 length was observed, further reducing this length can also show improvement in topics.

THANK YOU!!

[NoteBook Link](#)