Computational Text Analysis for Content Analysis

By Vaishali Kushwaha and Tieanna Graphenreed
Digital Integration Teaching Initiative (DITI)

POLS 2395 Environmental Politics

Daniel Aldrich

Spring 2022

Northeastern University
NULab for Texts, Maps, and Networks

Feel free to ask questions at any point during the presentation!

Workshop Agenda

- Introduction to definitions and key terms in computational text analysis (What)
- Discussion on its applications and uses in research (Why)
- Demonstration of web-based text analysis tools (How)
    - Word Counter, Word Trees, Voyant, Lexos

Slides, handouts, and data available at

http://bit.ly/diti-spring2022-Env_Pol-aldrich

Feel free to ask questions at any point during the presentation!

Workshop Outline

- Introduction
- Examples from practice
- Text preparation
- Tools:
  - Word Counter and Word Trees
  - Voyant
  - Lexos

Feel free to ask questions at any point during the presentation!

Introduction

Computational Text Analysis

Text analysis is a process to make inferences based on textual data. Computational text analysis refers to the array of methods used to "read" texts with a computer. It is similar to statistical analysis, but the data are texts (words) instead of numbers.

Text analysis:

- Involves a computer drawing out patterns in a text, and a researcher interpreting those patterns.
- Includes methods such as word count frequency, nGrams, and sentiment analysis.

Feel free to ask questions at any point during the presentation!

Why Computational Text Analysis?

Computational text analysis can help us **analyze very large amounts of data, identify keywords**, and **discover patterns** in a corpus of text(s). Using text analysis, researchers may **find surprising results** that they would not have discovered from close reading or traditional methods alone.

From collections of texts, researchers can discover keywords that serve as a proxy for major trends in societies, cultures, and policies. For example, computational tools can reveal patterns on how public officials communicate policies, which issues are of concern, which phrases leaders regularly employ, and much more.

Feel free to ask questions at any point during the presentation!

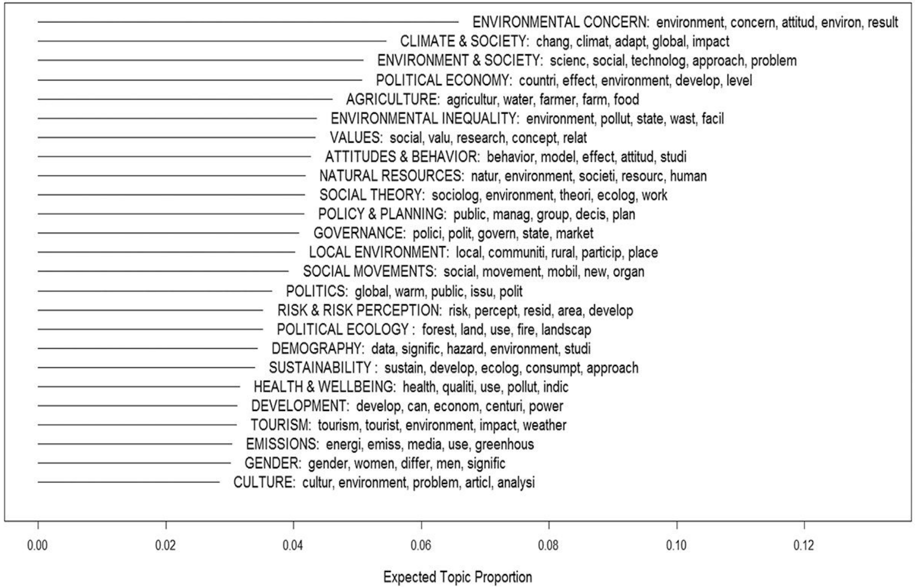
Key Terms

- **Corpus (plural–corpora)**: A collection of texts used for analysis and research purposes.
- **Stop words**: Words that appear frequently in a language, like pronouns, prepositions, and basic verbs. These are often removed for computational analysis. Some English stop words include: a, the, she, he, I, me, us, of, is, would, could, should, etc.
- **Word Count Frequency:** Counting the total times a word appears in a text/corpus or the percentage of how often it appears.
- **nGram:** A continuous sequence of *n* items in a text. A bigram (or 2 continuous words) could be ‘United States,’ while a trigram (3 words) could be ‘yes we can.’
- **Sentiment Analysis**: Measuring the sentiment of a text based on a scale such as negative/positive or happy/sad. Each word has a particular weight to determine where on the scale it falls, and these weights are calculated to determine a text’s overall sentiment.

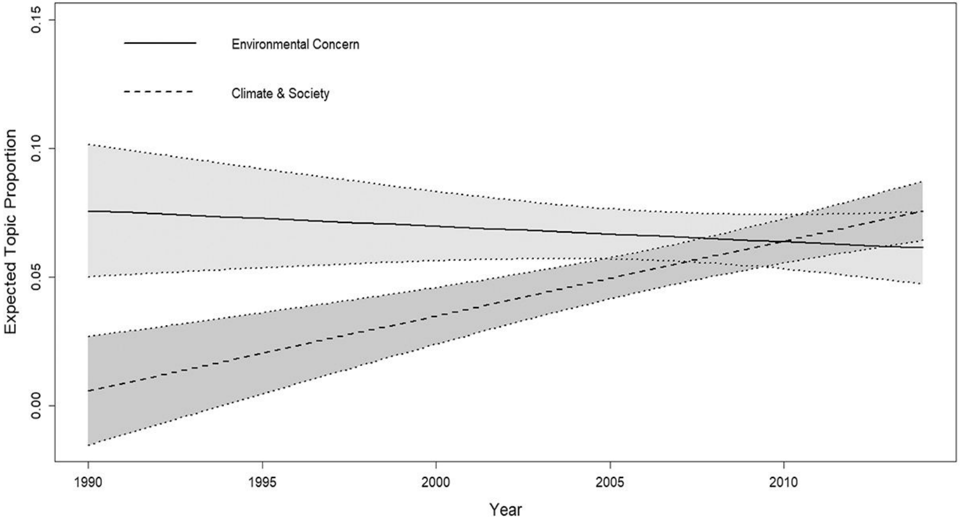
Examples from Practice

Feel free to ask questions at any point during the presentation!

Key Topics in Environmental Sociology



25 topics ranked from most to least prevalent in the corpus of 815 environmental sociology articles, including the top five associated word stems. The *x*-axis represents the proportion of each topic within the overall corpus.



Topical prevalence of 'Environmental Concern' and 'Climate & Society' over time, 1990 – 2014 (with 95% confidence intervals).

Jeremiah Bohr & Riley E. Dunlap (2018) Key Topics in environmental sociology, 1990–2014: results from a computational text analysis, Environmental Sociology, 4:2, 181-195, DOI: 10.1080/23251042.2017.1393863

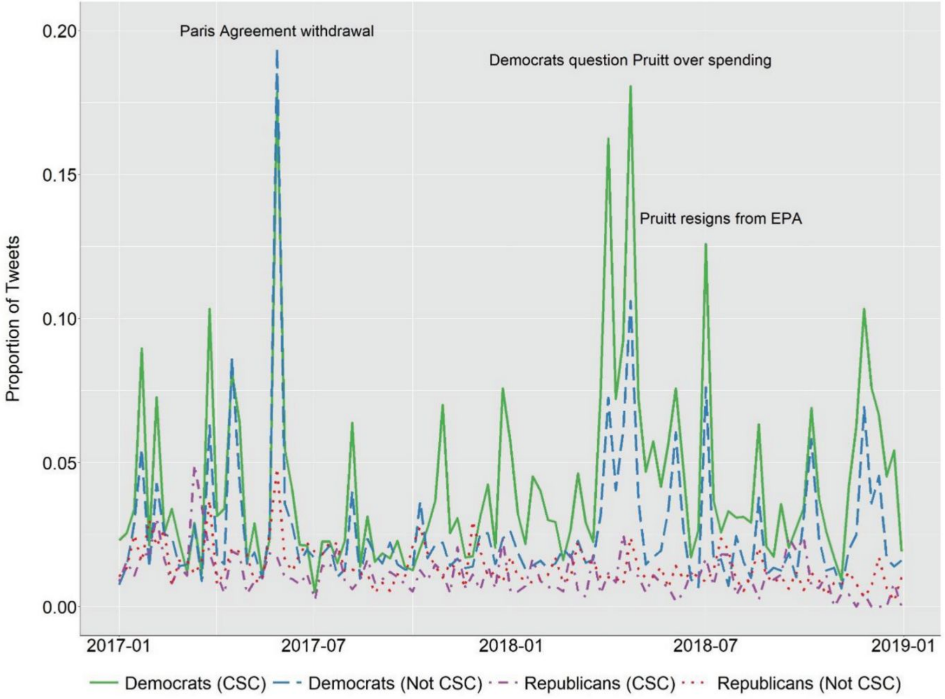
Northeastern University
NULab for Texts, Maps, and Networks

Feel free to ask questions at any point during the presentation!

U.S. Environmental Politics

To what extent politicians publicly discuss environmental issues in line with public opinion and economic characteristics of their constituents?

- Nominally pro-environment Republicans representing more moderate constituents fail to oppose their partisan colleagues, particularly during the Trump administration's withdrawal from the Paris Agreement. At the same time, very few openly attacked climate science



Weekly proportions of tweets discussing environmental issues sent by the 115th House of Representatives.

Key events and challenges: a computational text analysis of the 115th house of representatives on Twitter - Jeremiah Bohr in Environmental Politics (2021), 30 (3): 399-422

Additional Examples

- National interests and coalition positions on climate change: A text-based analysis - Paula Castro in *International Political Science Review* (2020) ,42 (1): 95-113
- The Meaning of Action: Linking Goal Orientations, Tactics, and Strategies in the Environmental Movement - Laura K. Nelson and Brayden G King in *Mobilization: An International Quarterly* (2020) 25 (3): 315–338.
- Posts mentioning ‘Black lives matter’ spiked on lawmakers’ social media accounts after the death of George Floyd - Pew Research Centre

Text Preparation

Corpus Building

Questions to consider before you begin:

- What are my research questions and why am I creating a corpus?
- What am I asking my corpus to do?
- What text(s) should form my corpus to answer my research questions?
- How should I organize my corpus to streamline my research processes and save time?

Feel free to ask questions at any point during the presentation!

Preparing Your Text

1. Choose the texts or text selections that you would like to include.
2. Create a folder on your computer or cloud storage where you will store your corpus. Give it a clearly descriptive name, without spaces or special characters.
3. If you are using a text that isn't already plain text, then copy and paste your text into a **plain text editor** (on Macs: Text Edit; on Windows: Notepad)
    a. Mac users, you may need to make your Text Edit into a 'plain text'. Open Text Edit, go to Preferences, and make sure "plain text" is selected
4. Save the text as a plain text file (with a .txt extension). Always make sure to name your files so you know what is in them!
5. Repeat steps above for each text in the corpus.

Feel free to ask questions at any point during the presentation!

Our Text

Our text is plain text (.txt file) of President Joe Biden's speech at COP26 Climate Summit 2021 at Glasgow. The primary objective is to explore this text using web-based computational text analysis tools.

We will also use Greta Thunberg's and Kausea Natano, Prime Minister of Tuvalu to see how a corpus can be analyzed. The primary objective is to compare and contrast the three speeches.

(Tuvalu: An island nation in Oceania)

Sample Corpus

The following .txt files are available on:

http://bit.ly/diti-spring2022-aldrich

Feel free to ask questions at any point during the presentation!

Exploratory Tool: Word Counter

Northeastern University
NULab for Texts, Maps, and Networks

Feel free to ask questions at any point during the presentation!

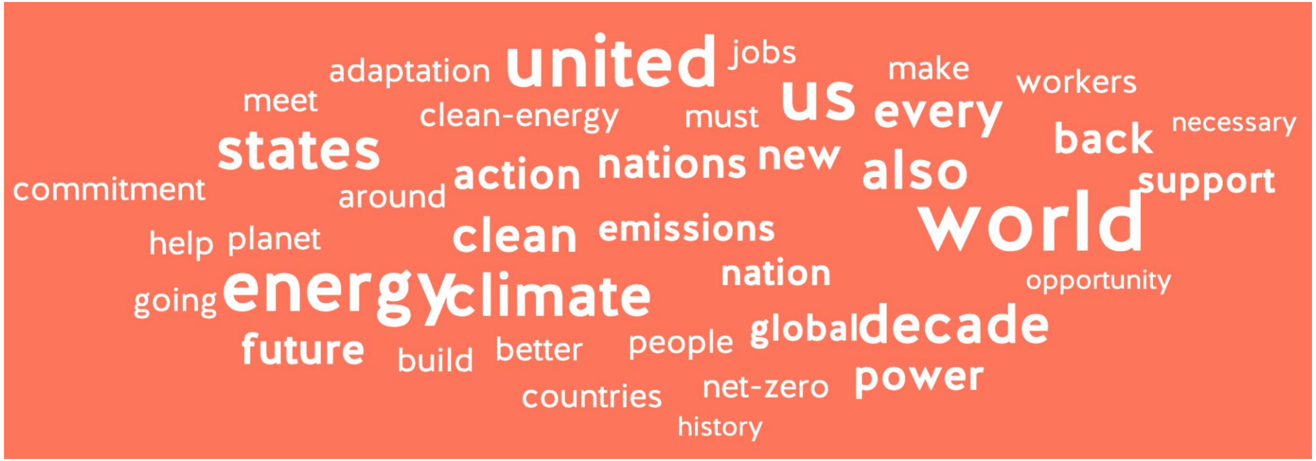
Word Counter

- https://databasic.io/en/wordcounter/
- A user-friendly **basic word counting tool**
- It allows you to count words, bigrams, and trigrams in plain text files and to download spreadsheets with your results
- The max file upload is 10MB
- The default is to lowercase all words and remove stopwords, but you can control these options

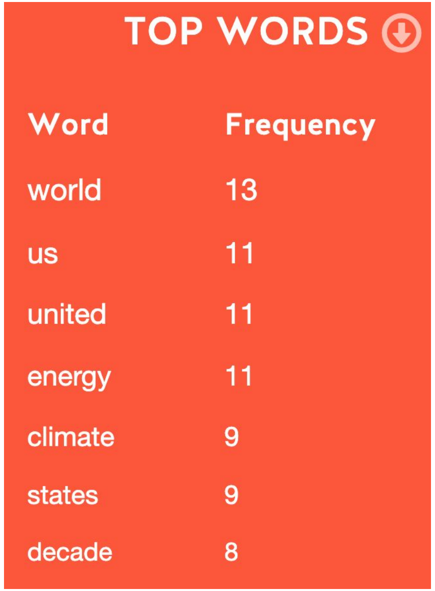
Feel free to ask questions at any point during the presentation!

Word Counter Examples

Word Counter will show you a word cloud, which can give you a sense of the **most used words in a document**. Words used more often are bigger, and ones used less often are smaller.



Word Counter Examples

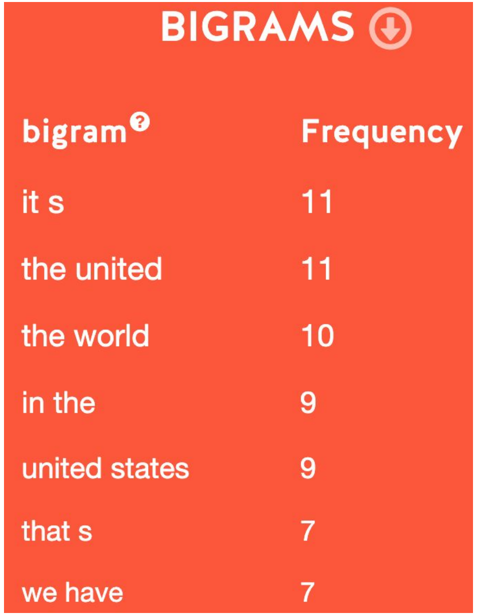


TOP WORDS

Word	Frequency
world	13
us	11
united	11
energy	11
climate	9
states	9
decade	8

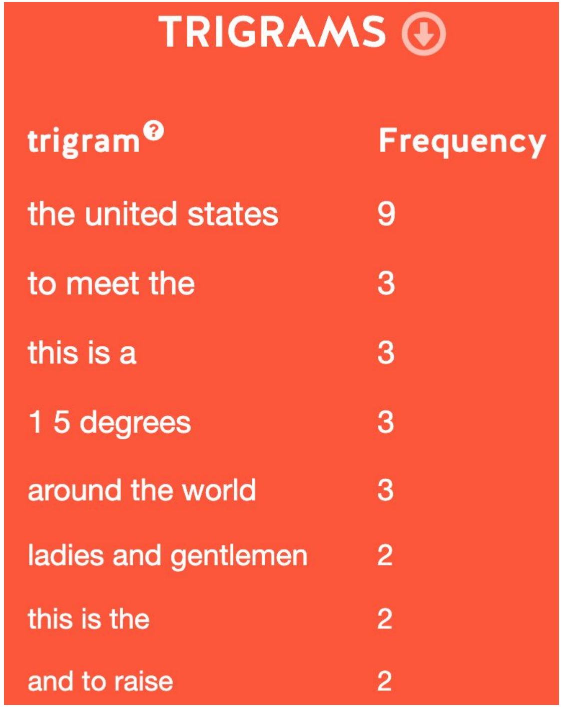
Shows the top words in the text.

Stopwords aren't removed for the bigrams and trigrams because they need context.



BIGRAMS

bigram	Frequency
it s	11
the united	11
the world	10
in the	9
united states	9
that s	7
we have	7



TRIGRAMS

trigram	Frequency
the united states	9
to meet the	3
this is a	3
1 5 degrees	3
around the world	3
ladies and gentlemen	2
this is the	2
and to raise	2

Feel free to ask questions at any point during the presentation!

Exploratory Tool:
Word Tree

Northeastern University
NULab for Texts, Maps, and Networks

Feel free to ask questions at any point during the presentation!

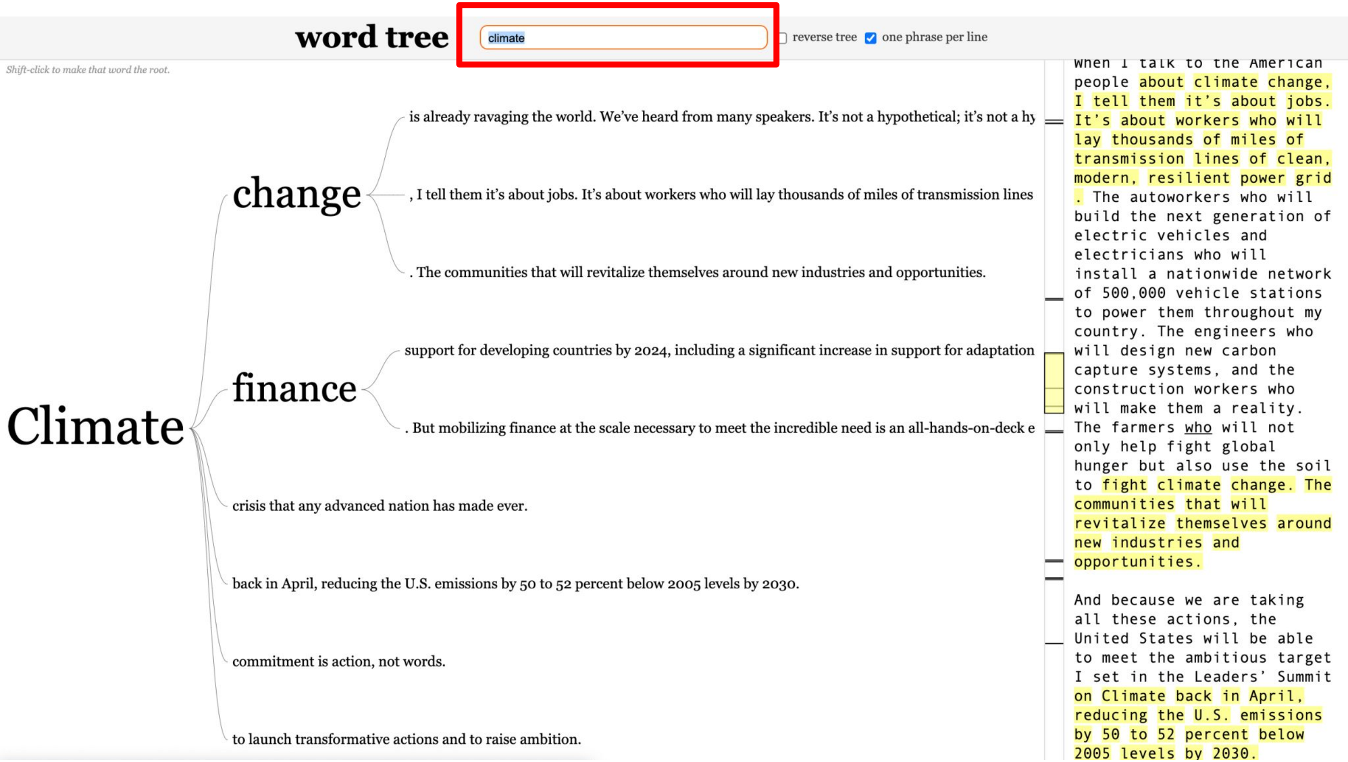
Word Tree

- https://www.jasondavies.com/wordtree/
- A word tree **depicts multiple parallel sequences of words**
- This is a good way to see patterns in word usage, based on words that appear before and after a term or terms of interest.
- There are some restrictions in size: fewer than 1 million words should work

Feel free to ask questions at any point during the presentation!

Word Tree Example

Reflects the focus of the speech on climate change and finance.



word tree climate □ reverse tree ☑ one phrase per line

Shift-click to make that word the root.

Climate

- change
  - is already ravaging the world. We've heard from many speakers. It's not a hypothetical; it's not a hy
  - , I tell them it's about jobs. It's about workers who will lay thousands of miles of transmission lines
  - . The communities that will revitalize themselves around new industries and opportunities.
- finance
  - support for developing countries by 2024, including a significant increase in support for adaptation
  - . But mobilizing finance at the scale necessary to meet the incredible need is an all-hands-on-deck e
- crisis that any advanced nation has made ever.
- back in April, reducing the U.S. emissions by 50 to 52 percent below 2005 levels by 2030.
- commitment is action, not words.
- to launch transformative actions and to raise ambition.

```
when I talk to the American
people about climate change,
I tell them it's about jobs.
It's about workers who will
lay thousands of miles of
transmission lines of clean,
modern, resilient power grid
. The autoworkers who will
build the next generation of
electric vehicles and
electricians who will
install a nationwide network
of 500,000 vehicle stations
to power them throughout my
country. The engineers who
will design new carbon
capture systems, and the
construction workers who
will make them a reality.
The farmers who will not
only help fight global
hunger but also use the soil
to fight climate change. The
communities that will
revitalize themselves around
new industries and
opportunities.

And because we are taking
all these actions, the
United States will be able
to meet the ambitious target
I set in the Leaders' Summit
on Climate back in April,
reducing the U.S. emissions
by 50 to 52 percent below
2005 levels by 2030.
```

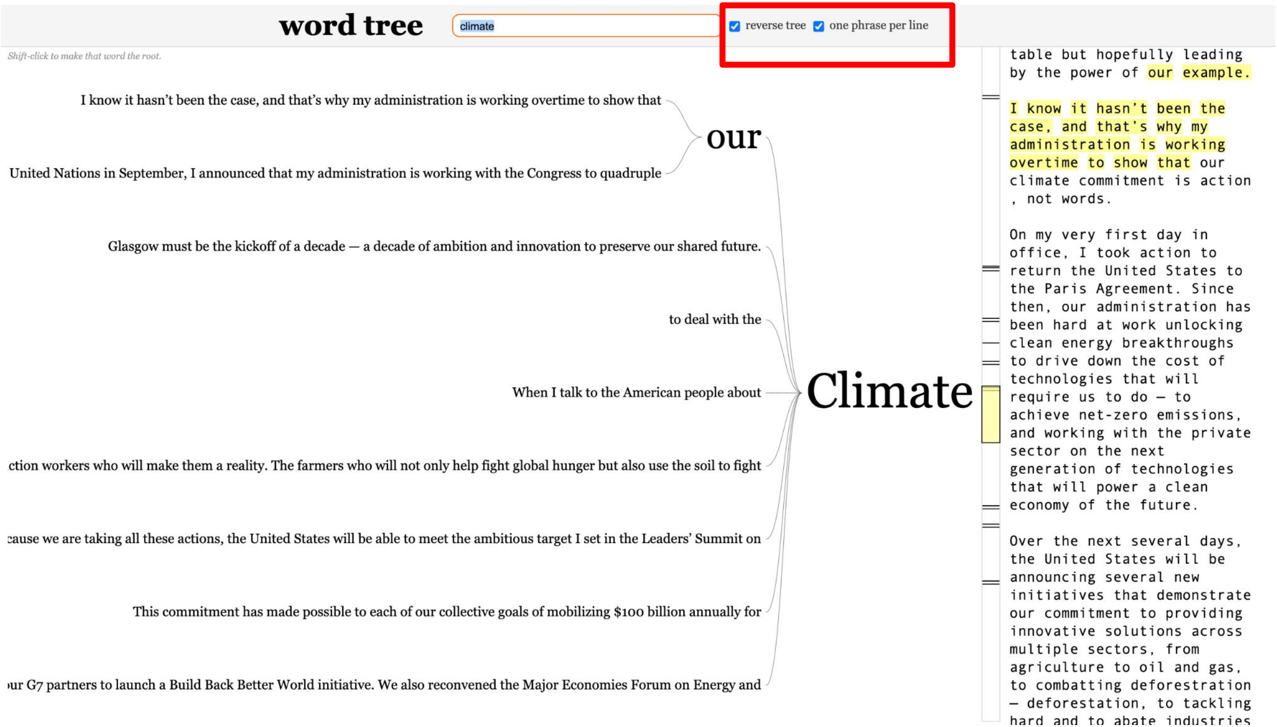
Northeastern University

NULab for Texts, Maps, and Networks

Feel free to ask questions at any point during the presentation!

Word Tree: Reverse Trees

It is worth reversing the tree to see the words that often precede it. To do this click “reverse tree” next to the search bar.



Shift-click to make that word the root.

I know it hasn’t been the case, and that’s why my administration is working overtime to show that our

United Nations in September, I announced that my administration is working with the Congress to quadruple our

Glasgow must be the kickoff of a decade — a decade of ambition and innovation to preserve our shared future.

to deal with the

When I talk to the American people about

ction workers who will make them a reality. The farmers who will not only help fight global hunger but also use the soil to fight

cause we are taking all these actions, the United States will be able to meet the ambitious target I set in the Leaders’ Summit on

This commitment has made possible to each of our collective goals of mobilizing $100 billion annually for

ur G7 partners to launch a Build Back Better World initiative. We also reconvened the Major Economies Forum on Energy and

Climate

```
table but hopefully leading
by the power of our example.

I know it hasn’t been the
case, and that’s why my
administration is working
overtime to show that our
climate commitment is action
, not words.

On my very first day in
office, I took action to
return the United States to
the Paris Agreement. Since
then, our administration has
been hard at work unlocking
clean energy breakthroughs
to drive down the cost of
technologies that will
require us to do — to
achieve net-zero emissions,
and working with the private
sector on the next
generation of technologies
that will power a clean
economy of the future.

Over the next several days,
the United States will be
announcing several new
initiatives that demonstrate
our commitment to providing
innovative solutions across
multiple sectors, from
agriculture to oil and gas,
to combatting deforestation
— deforestation, to tackling
hard and to abate industries
```

Your Turn!

Use the sample text or texts of your choice and begin practicing web-browser text analysis. **Explore Word Counter and Word Tree!**

Discussion Prompts

- What limitations are you observing? What functionalities do you wish these tools might offer?
- Even with these limitations, how can you apply these simple tools in your research and exploration?

Feel free to ask questions at any point during the presentation!

Powerful Platform: Voyant

Northeastern University
NULab for Texts, Maps, and Networks

Feel free to ask questions at any point during the presentation!

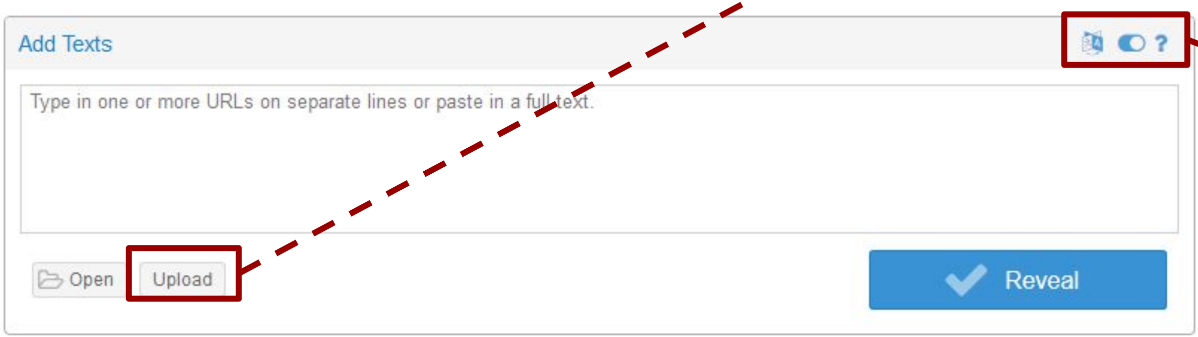
Voyant

Voyant makes it possible to **perform analyses on one or multiple files in many ways**, including word counts, nGrams (n=number of words), word frequency distributions, word trends across documents, and concordances. It also makes nice visualizations!

https://voyant-tools.org/

VOYANT

see through your text



Type in one or more URLs on separate lines or paste in a full text.

Click on Upload and navigate to the folder with the text document you wish to analyze.

Alternatively, insert URLs or full text into the textbox.

Click here for help and advanced options

Northeastern University
NULab for Texts, Maps, and Networks

Feel free to ask questions at any point during the presentation!

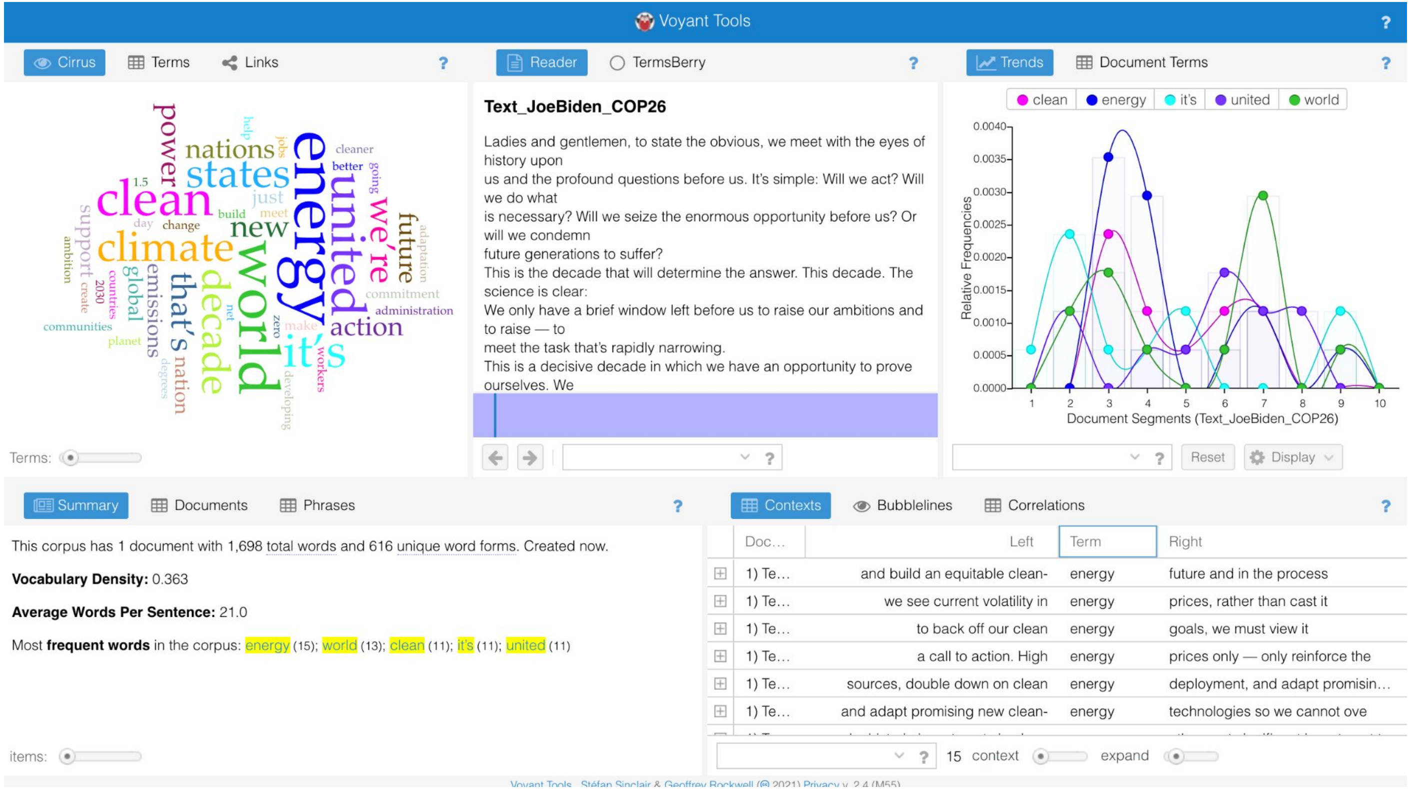
Voyant: Basic Dashboard

Results:

From Climate Ready Boston you can see the default results page with multiple panes:

- A word cloud
- Reader section
- Trends
- Document summary
- Word contexts

These boxes can all be changed!



Text_JoeBiden_COP26

Ladies and gentlemen, to state the obvious, we meet with the eyes of history upon
us and the profound questions before us. It's simple: Will we act? Will we do what
is necessary? Will we seize the enormous opportunity before us? Or will we condemn
future generations to suffer?
This is the decade that will determine the answer. This decade. The science is clear:
We only have a brief window left before us to raise our ambitions and to raise — to
meet the task that's rapidly narrowing.
This is a decisive decade in which we have an opportunity to prove ourselves. We

This corpus has 1 document with 1,698 total words and 616 unique word forms. Created now.

Vocabulary Density: 0.363

Average Words Per Sentence: 21.0

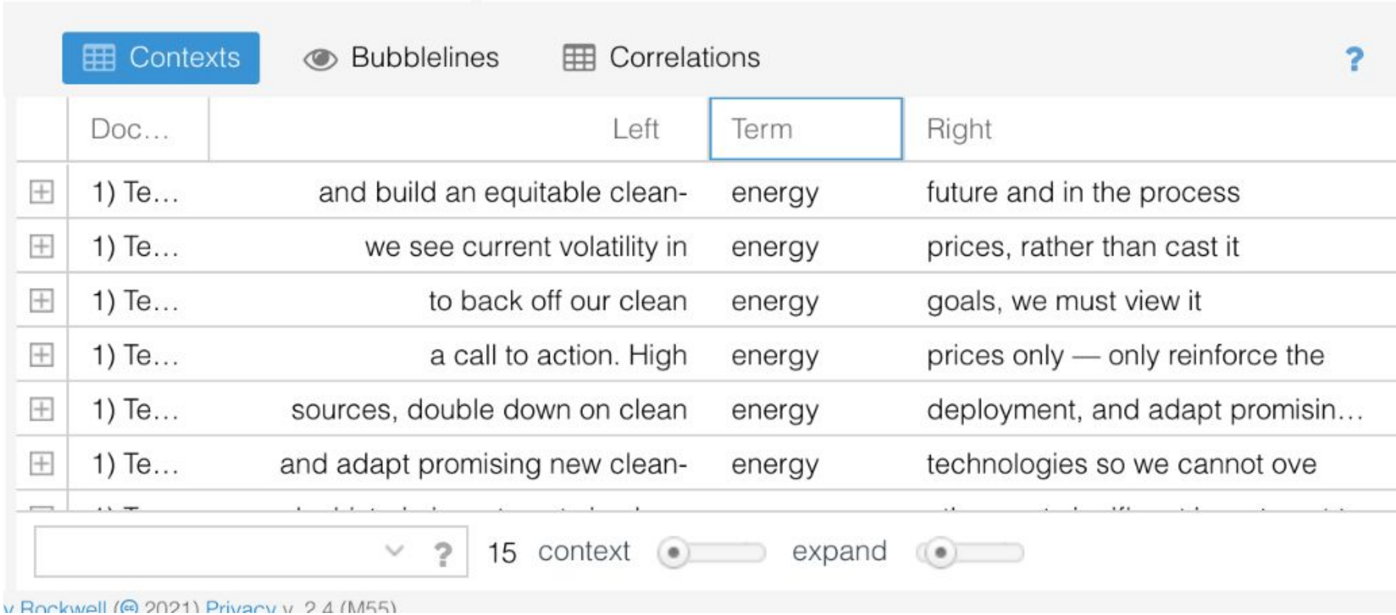
Most **frequent words** in the corpus: energy (15); world (13); clean (11); it's (11); united (11)

	Doc…	Left	Term	Right
⊞	1) Te…	and build an equitable clean-	energy	future and in the process
⊞	1) Te…	we see current volatility in	energy	prices, rather than cast it
⊞	1) Te…	to back off our clean	energy	goals, we must view it
⊞	1) Te…	a call to action. High	energy	prices only — only reinforce the
⊞	1) Te…	sources, double down on clean	energy	deployment, and adapt promisin…
⊞	1) Te…	and adapt promising new clean-	energy	technologies so we cannot ove

Voyant Tools , Stéfan Sinclair & Geoffrey Rockwell (© 2021) Privacy v. 2.4 (M55)

Voyant: Contexts (concordances)

Contexts, or concordances, show the different contexts around particular search terms. For example, you can see all the times the word “energy” appears in the text and the contexts in which it appears.

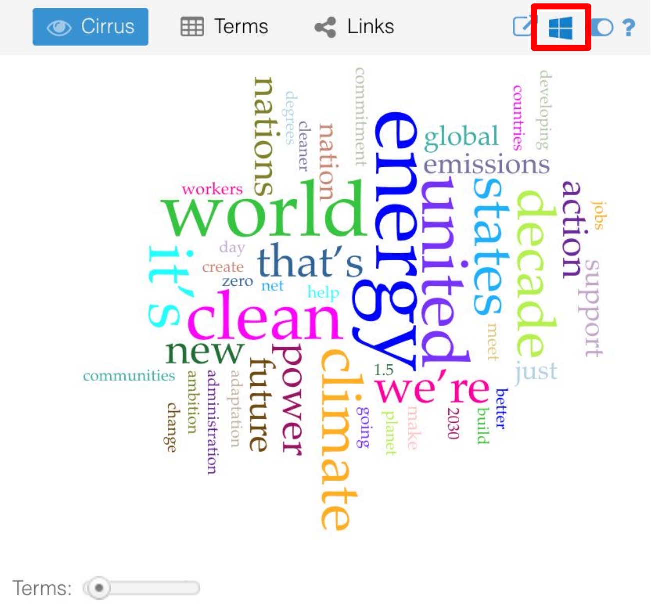


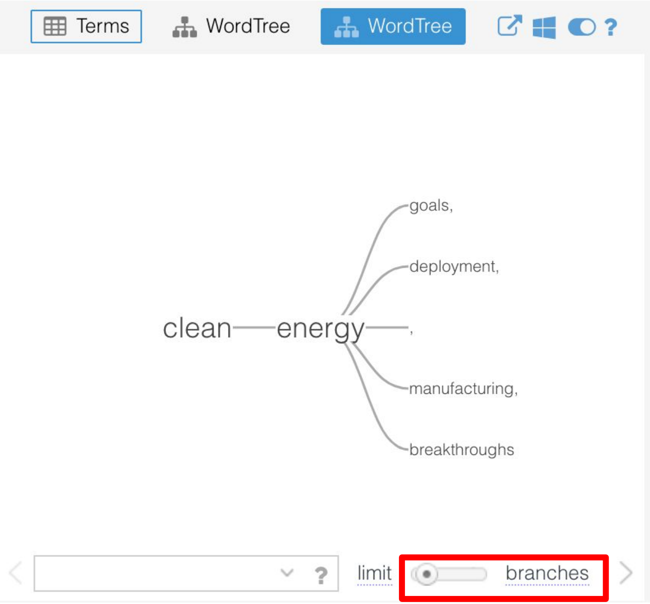
	Doc…	Left	Term	Right
⊞	1) Te…	and build an equitable clean-	energy	future and in the process
⊞	1) Te…	we see current volatility in	energy	prices, rather than cast it
⊞	1) Te…	to back off our clean	energy	goals, we must view it
⊞	1) Te…	a call to action. High	energy	prices only — only reinforce the
⊞	1) Te…	sources, double down on clean	energy	deployment, and adapt promisin…
⊞	1) Te…	and adapt promising new clean-	energy	technologies so we cannot ove

v Rockwell (© 2021) Privacy v. 2.4 (M55)

Voyant: Changing displayed results

Hover on the right top corner of a pane and buttons will appear. Select the panes button and choose a new option from the dropdown menu





For our new pane option, we have chosen the WordTree visualization from the 'visualization tools' dropdown sub-menu. You can select the number of "branches" by dragging the scroll button at the bottom.

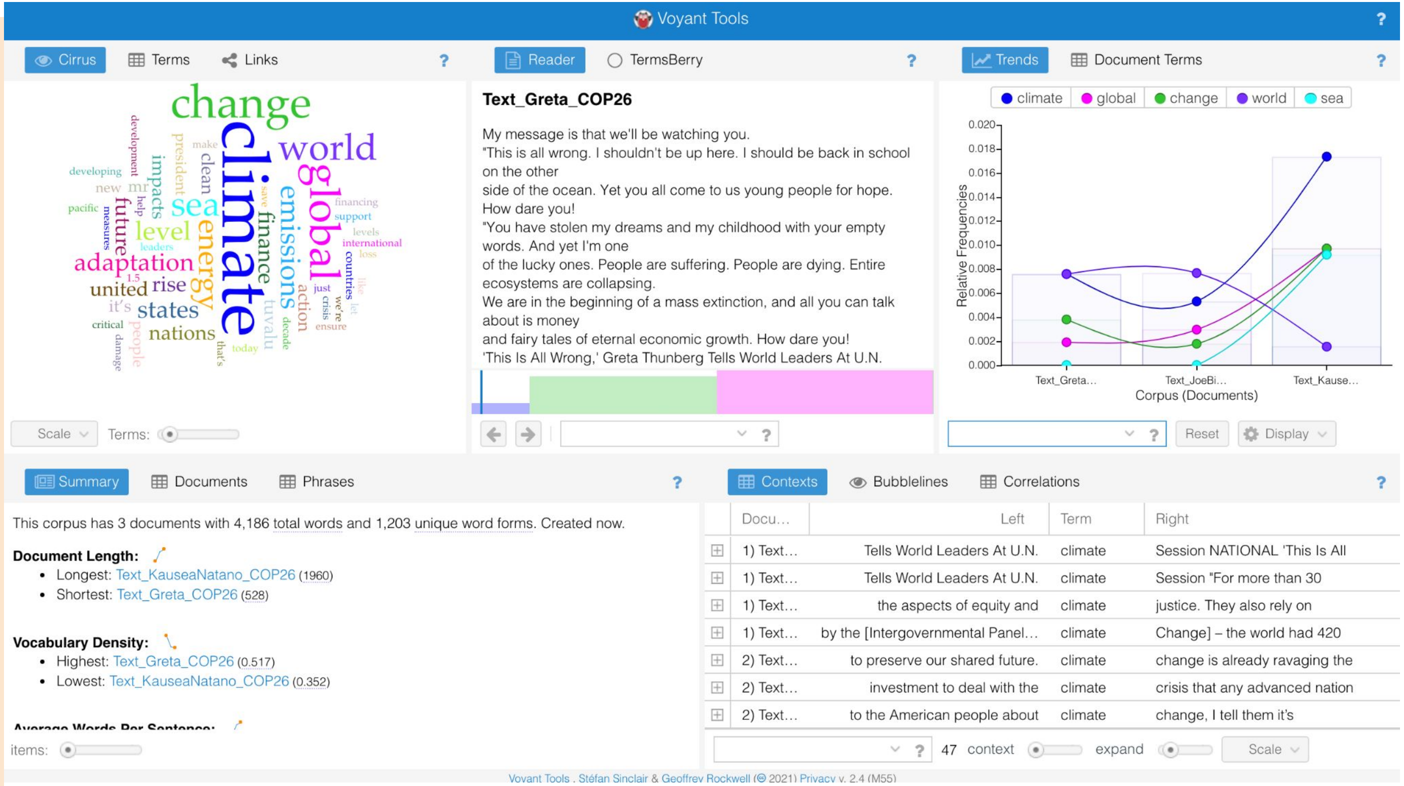
Northeastern University
NULab for Texts, Maps, and Networks

Feel free to ask questions at any point during the presentation!

Voyant: Corpus Dashboard

Results page of the corpus containing climate reports of 5 cities.

- A word cloud: combining all texts
- Reader section: scroll down all texts
- **Trends: relative frequency of terms across all texts - good for comparison**
- **Document Summary- good for comparison**
- Word Contexts: separate for all texts



Text_Greta_COP26

My message is that we'll be watching you.
"This is all wrong. I shouldn't be up here. I should be back in school on the other
side of the ocean. Yet you all come to us young people for hope.
How dare you!
"You have stolen my dreams and my childhood with your empty words. And yet I'm one
of the lucky ones. People are suffering. People are dying. Entire ecosystems are collapsing.
We are in the beginning of a mass extinction, and all you can talk about is money
and fairy tales of eternal economic growth. How dare you!
'This Is All Wrong,' Greta Thunberg Tells World Leaders At U.N.

This corpus has 3 documents with 4,186 total words and 1,203 unique word forms. Created now.

Document Length:

- Longest: Text_KauseaNatano_COP26 (1960)
- Shortest: Text_Greta_COP26 (528)

Vocabulary Density:

- Highest: Text_Greta_COP26 (0.517)
- Lowest: Text_KauseaNatano_COP26 (0.352)

Average Words Per Sentence:

	Docu...	Left	Term	Right
⊞	1) Text...	Tells World Leaders At U.N.	climate	Session NATIONAL 'This Is All
⊞	1) Text...	Tells World Leaders At U.N.	climate	Session "For more than 30
⊞	1) Text...	the aspects of equity and	climate	justice. They also rely on
⊞	1) Text...	by the [Intergovernmental Panel...	climate	Change] – the world had 420
⊞	2) Text...	to preserve our shared future.	climate	change is already ravaging the
⊞	2) Text...	investment to deal with the	climate	crisis that any advanced nation
⊞	2) Text...	to the American people about	climate	change, I tell them it's

Voyant Tools . Stéfan Sinclair & Geoffrey Rockwell (© 2021) Privacy v. 2.4 (M55)

Northeastern University
NULab for Texts, Maps, and Networks

Feel free to ask questions at any point during the presentation!

Your Turn!

Use the sample text or texts of your choice and begin practicing web-browser text analysis. **Explore Voyant's features!**

Discussion Prompts

- What do you find challenging or exciting about this tool?
- What interesting or surprising results came up?
- How might you interpret those results based on what you know about your field?

Feel free to ask questions at any point during the presentation!

Powerful Platform: Lexos

Feel free to ask questions at any point during the presentation!

Lexos

Lexos provides a step-by-step guide for text uploading, preparation, and analysis.

- **Upload**: upload your .txt file
- **Manage**: select the files you want to prepare and analyze
- **Prepare**: prepare your text for analysis
- **Visualize**: create visualizations of patterns across your corpus or in single texts
- **Analyze**: analyze your text

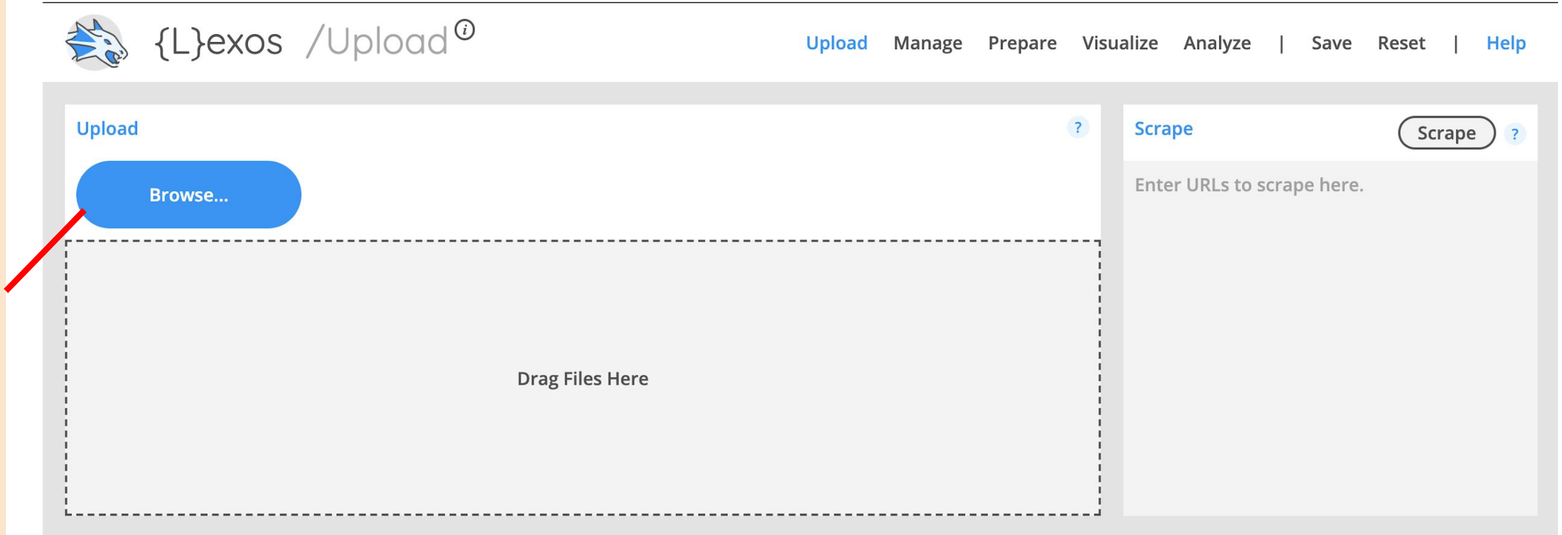
http://lexos.wheatoncollege.edu/upload

Feel free to ask questions at any point during the presentation!

Lexos: Upload

Click Browse and select your entire text (or drag file into the “Drag Files Here” area)

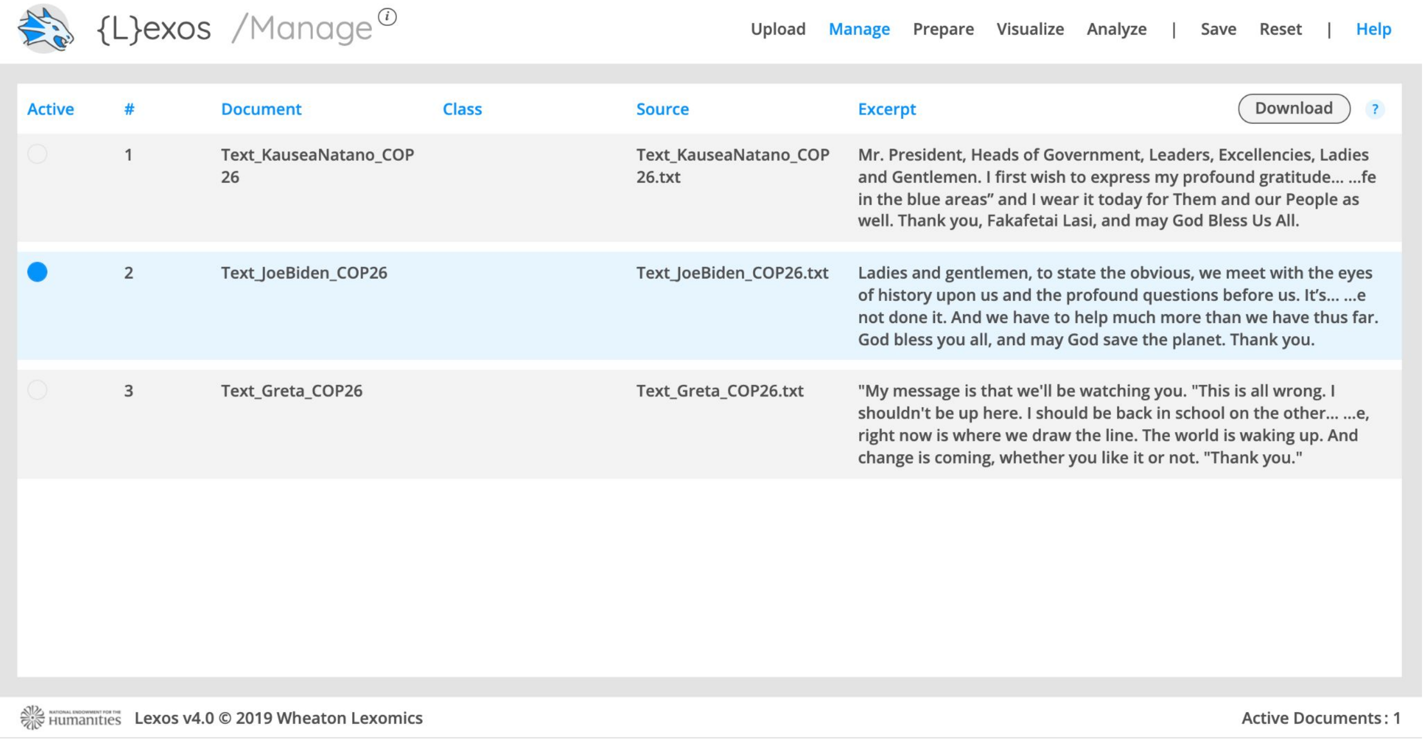
You will not get a super visible notification when the upload is done - click “Manage” to double check that the text file is there.



Feel free to ask questions at any point during the presentation!

Lexos: Manage

Make sure the document you want to use is selected (blue = selected, gray = not selected)



Active	#	Document	Class	Source	Excerpt
○	1	Text_KauseaNatano_COP26		Text_KauseaNatano_COP26.txt	Mr. President, Heads of Government, Leaders, Excellencies, Ladies and Gentlemen. I first wish to express my profound gratitude… …fe in the blue areas" and I wear it today for Them and our People as well. Thank you, Fakafetai Lasi, and may God Bless Us All.
●	2	Text_JoeBiden_COP26		Text_JoeBiden_COP26.txt	Ladies and gentlemen, to state the obvious, we meet with the eyes of history upon us and the profound questions before us. It's… …e not done it. And we have to help much more than we have thus far. God bless you all, and may God save the planet. Thank you.
○	3	Text_Greta_COP26		Text_Greta_COP26.txt	"My message is that we'll be watching you. "This is all wrong. I shouldn't be up here. I should be back in school on the other… …e, right now is where we draw the line. The world is waking up. And change is coming, whether you like it or not. "Thank you."

Lexos v4.0 © 2019 Wheaton Lexomics

Active Documents : 1

Northeastern University
NULab for Texts, Maps, and Networks

Feel free to ask questions at any point during the presentation!

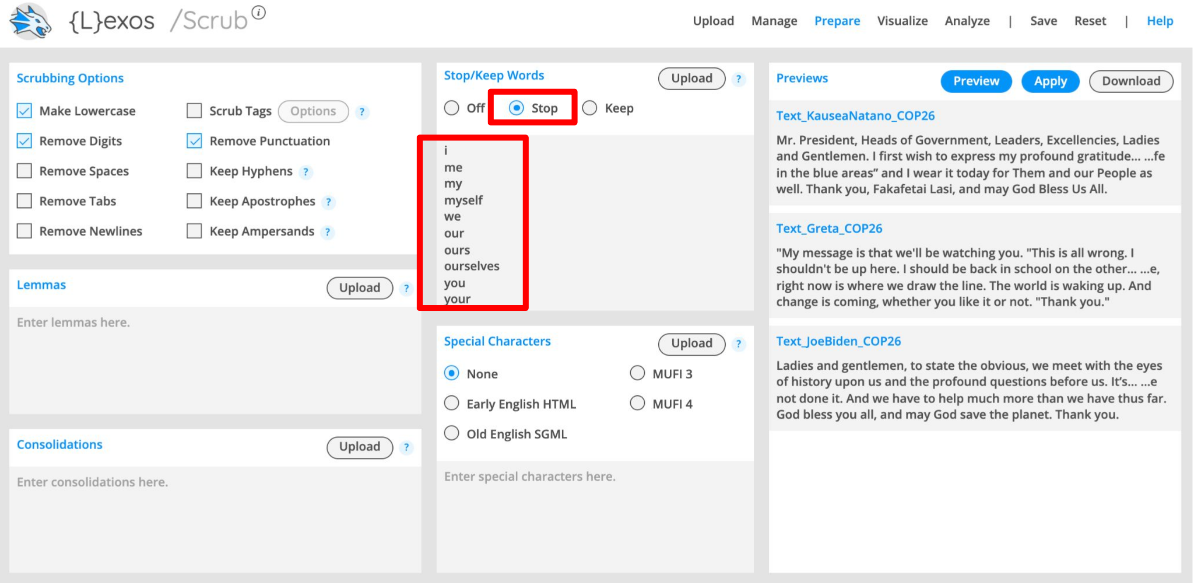
Lexos: Prepare (scrub)

Lexos demonstrates some more advanced options you have for preparing your corpus. By “scrubbing,” you are transforming the texts in your corpus and making choices that will impact your results. Here are some possibilities:

- **Make Lowercase**: make all your letters lowercase. Even though you know “A” and “a” are the same letter, the computer treats these as two separate characters. Lowercasing removes this distinction.
- **Remove Punctuation**: remove punctuation, which may influence your results.
- **Stop/Keep Words**: remove a list of words (or keep only words from a list). Usually you would remove **stopwords**, or the most common words in a language (English: the, a she, her, it, him, they, etc).
- **Lemmas**: standardize to the *stem* of word. For example, you can stem all forms of talk: talking, talked, talks, etc. to “talk”

Lexos: Removing Stopwords

Get a list of English stopwords here: https://gist.github.com/sebleier/554280 (there is also a copy on the GitHub page). Copy and paste the stopwords (or upload the .txt file) into the "Stop/Keep Words" box then select "Stop"



Scrubbing Options

☑ Make Lowercase ☐ Scrub Tags Options ?

☑ Remove Digits ☑ Remove Punctuation

☐ Remove Spaces ☐ Keep Hyphens ?

☐ Remove Tabs ☐ Keep Apostrophes ?

☐ Remove Newlines ☐ Keep Ampersands ?

Lemmas Upload ?

Enter lemmas here.

Consolidations Upload ?

Enter consolidations here.

Stop/Keep Words Upload ?

○ Off ● Stop ○ Keep

i
me
my
myself
we
our
ours
ourselves
you
your

Special Characters Upload ?

● None ○ MUFI 3

○ Early English HTML ○ MUFI 4

○ Old English SGML

Enter special characters here.

Previews Preview Apply Download

Text_KauseaNatano_COP26

Mr. President, Heads of Government, Leaders, Excellencies, Ladies and Gentlemen. I first wish to express my profound gratitude... ...fe in the blue areas" and I wear it today for Them and our People as well. Thank you, Fakafetai Lasi, and may God Bless Us All.

Text_Greta_COP26

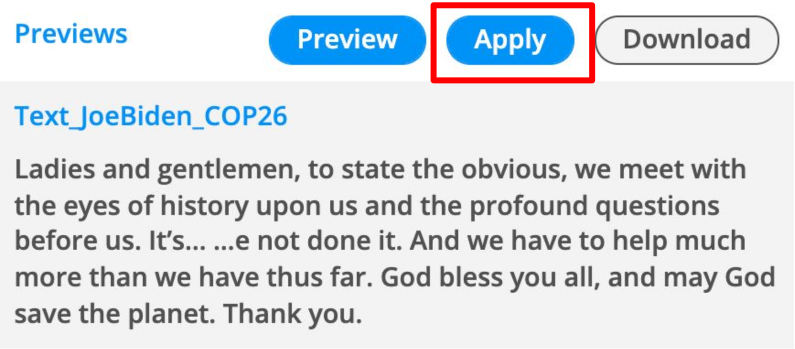
"My message is that we'll be watching you. "This is all wrong. I shouldn't be up here. I should be back in school on the other... ...e, right now is where we draw the line. The world is waking up. And change is coming, whether you like it or not. "Thank you."

Text_JoeBiden_COP26

Ladies and gentlemen, to state the obvious, we meet with the eyes of history upon us and the profound questions before us. It's... ...e not done it. And we have to help much more than we have thus far. God bless you all, and may God save the planet. Thank you.

Lexos: Applying your Preparations

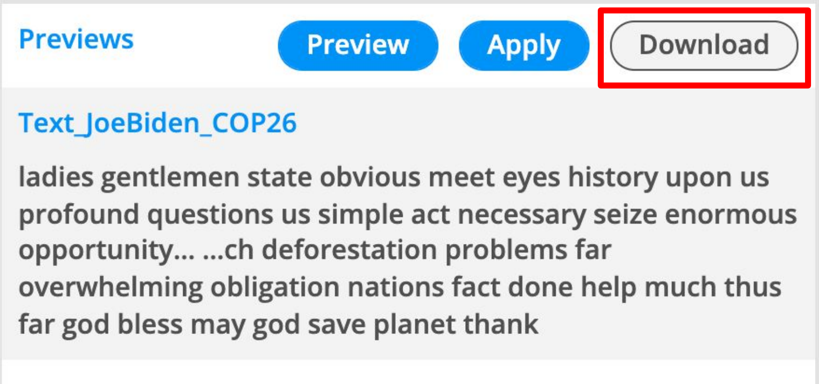
BEFORE PREP



Text_JoeBiden_COP26

Ladies and gentlemen, to state the obvious, we meet with the eyes of history upon us and the profound questions before us. It's… …e not done it. And we have to help much more than we have thus far. God bless you all, and may God save the planet. Thank you.

AFTER PREP



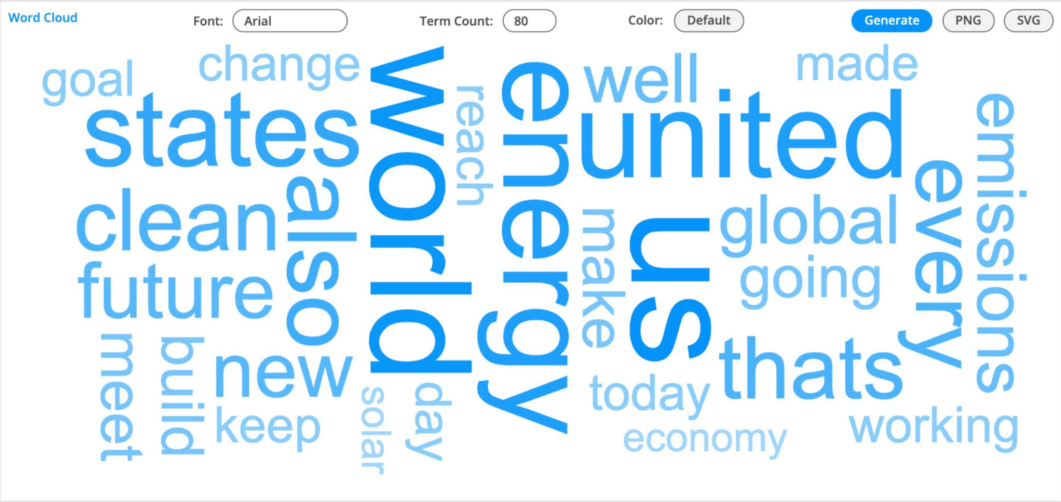
Text_JoeBiden_COP26

ladies gentlemen state obvious meet eyes history upon us profound questions us simple act necessary seize enormous opportunity… …ch deforestation problems far overwhelming obligation nations fact done help much thus far god bless may god save planet thank

Once you have made decisions about your preparations, click "**Apply**" and wait a few minutes. Because the program is going through each document and completing all the processes you selected, it needs some time. Then, you will see the final results of your preparation! You can also **download** your new corpus and use it with other tools.

Feel free to ask questions at any point during the presentation!

Lexos: Visualize



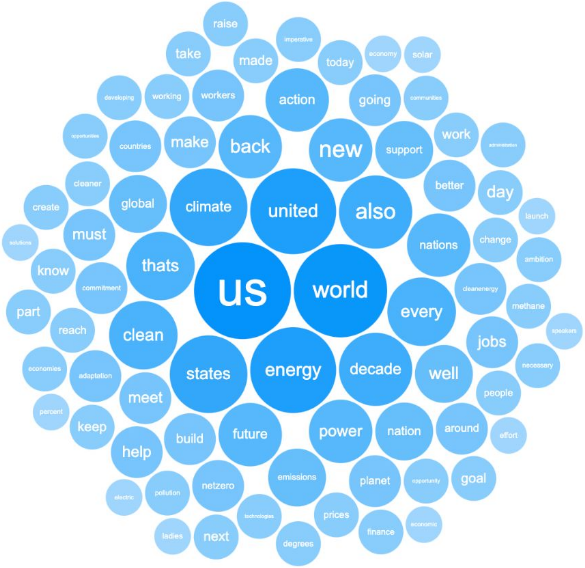
Font: Arial

Term Count: 80

Color: Default

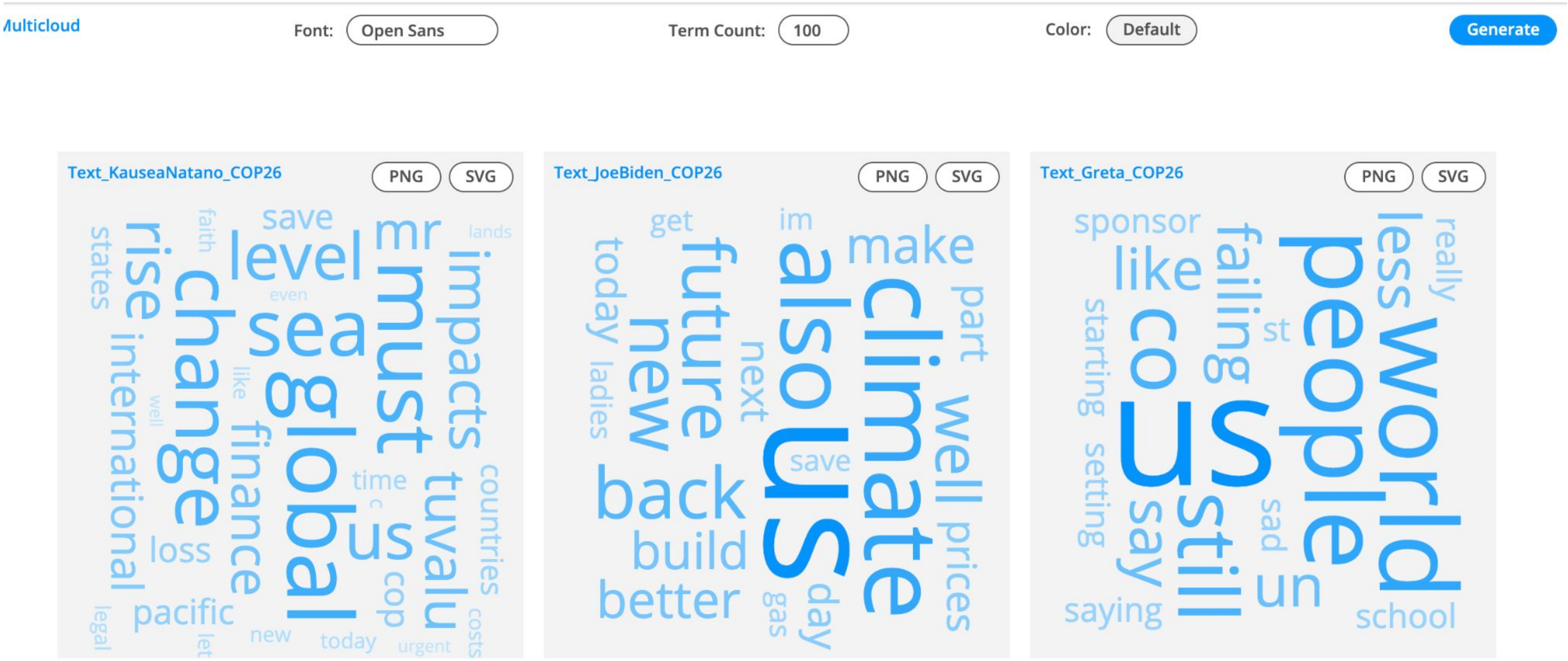
Word Cloud: visualize a wordcloud across the entire text/corpus.

Bubbleviz: visualize word counts through bubbles across the entire text/corpus.



Feel free to ask questions at any point during the presentation!

Lexos: Visualize > Multicloud



/ulticloud

Font: Open Sans

Term Count: 100

Color: Default

Generate

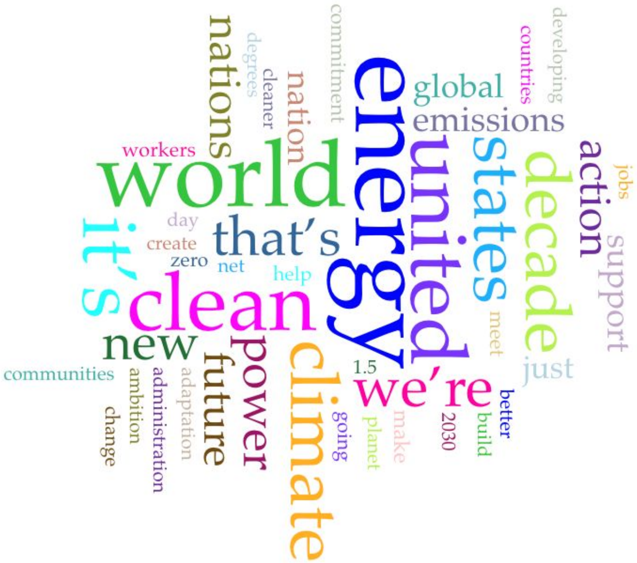
Text_KauseaNatano_COP26 PNG SVG

Text_JoeBiden_COP26 PNG SVG

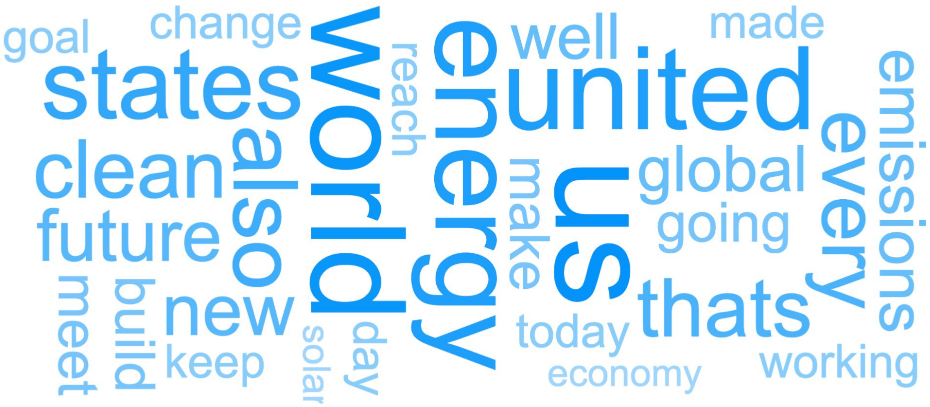
Text_Greta_COP26 PNG SVG

Voyant vs. Lexos: Wordclouds

How does the Voyant wordcloud below compare to the own made using Lexos?



Lexos Wordcloud



What could be causing this distinction? This helps demonstrate the importance of understanding what a tool is doing to the texts in the background.

Lexos: Rolling Window

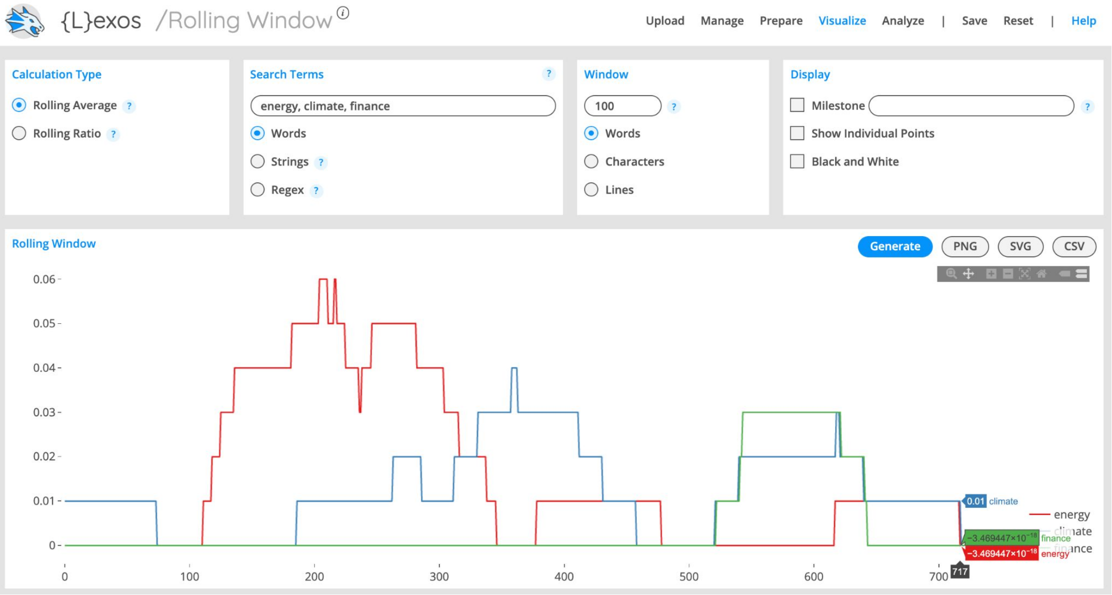
Rolling windows allow you to look at word trends across **one** document. To use a rolling window, first select a single text in the "Manage" screen, then:

1. Go to “Visualize-> Rolling Window” and type in a search term you want to visualize. You can also search multiple terms by clicking “String” and separating words with a comma (heat, health, flood, storm)
2. Choose a Window size (the number of words each “window” contains). For shorter documents, it’s good to have a number like 300/500. For larger documents, you may want to make your window larger. Play around with the window size until you get a visualization that makes sense.
3. Click “Generate”

Feel free to ask questions at any point during the presentation!

Lexos: Rolling Window Results

Using Joe Biden's speech, and searching for the words 'energy', 'climate' and 'finance' with a window of 100 (since this is a small document), we can get an idea of how these terms work together in the report.



Calculation Type

● Rolling Average ?

○ Rolling Ratio ?

Search Terms ?

energy, climate, finance

● Words

○ Strings ?

○ Regex ?

Window

100 ?

● Words

○ Characters

○ Lines

Display

□ Milestone ?

□ Show Individual Points

□ Black and White

Rolling Window

Feel free to ask questions at any point during the presentation!

Lexos: Analyze > Dendrogram

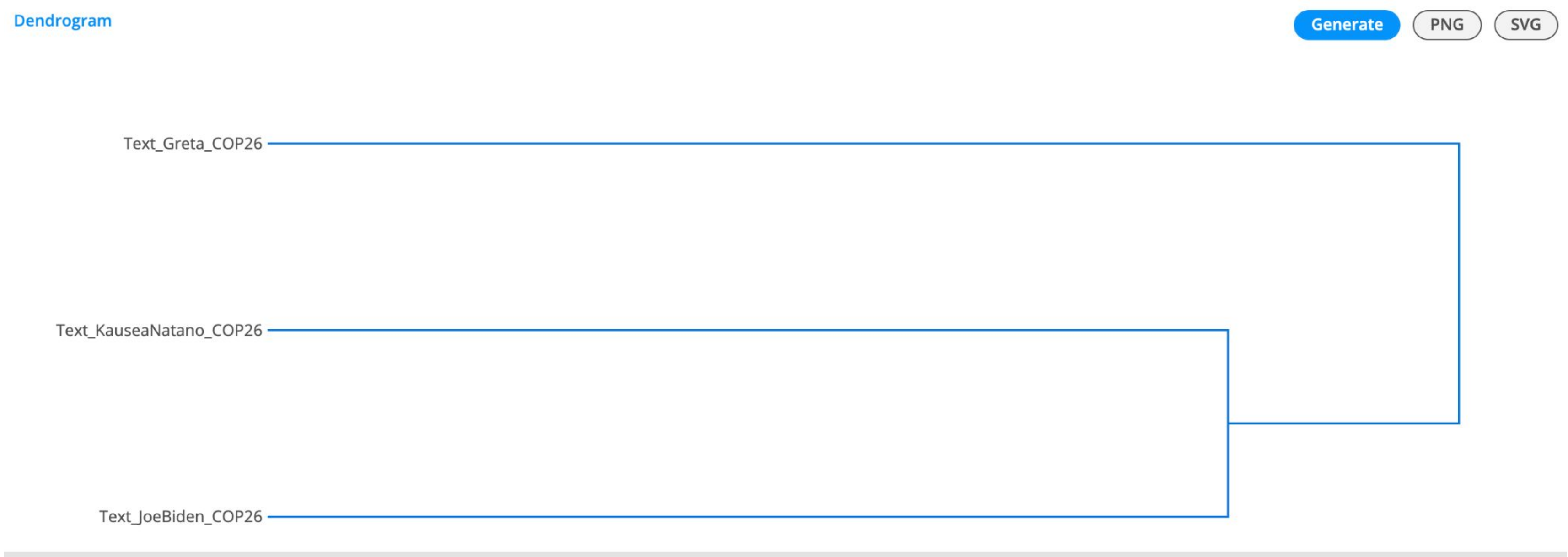
The dendrogram demonstrates similarity between the different documents. Dendrograms require at least two documents to compare. Dendrograms are able to show the hierarchy between objects. Dendrograms show:

- Similarities between texts
    - The greater the distance between texts, the less similar they are
    - The smaller the distance between texts, the more similar they are

Feel free to ask questions at any point during the presentation!

Lexos: Dendrogram

The dendrogram demonstrates similarity between the different documents.



Feel free to ask questions at any point during the presentation!

Lexos: Save or Reset Your Results

Lexos allows you to **save** your results as a Lexos file. If you do this, you can re-upload the Lexos file any time to access your cleaned-up corpus as well as the different analyses you've done. You can also download modified text files from the "Manage" page, which you can use with other tools if you would like.

You can also save individual visualizations as images (PNGs).

Finally, if you want to start over, you can "Reset" your Lexos dashboard.

Feel free to ask questions at any point during the presentation!

Your Turn!

Use the sample text or texts of your choice and begin practicing web-browser text analysis. **Explore Lexos's features!**

Discussion Prompts

- What difference did you notice between Voyant and Lexos?
- Which tool do you prefer and why?
- How would you want to use these tools in this class and future?

Thank you!

If you have any questions, contact us at nulab.info@gmail.com

Developed by Vaishali Kushwaha and Garrett Morrow
Delivered by Vaishali Kushwaha and Tieanna Graphenreed
DITI Research Fellows
Digital Integration Teaching Initiative

Slides, handouts, and data available at

http://bit.ly/diti-spring2022-Env_Pol-aldrich

You also have access to DITI Canvas Module on Computational Text Analysis.

Schedule an appointment with us! **https://calendly.com/diti-nu**

Feel free to ask questions at any point during the presentation!