



STEPHAN GRIMM

Tagging Podcasts

using Natural Language Processing (NLP)

Agenda

1 Task

2 Methodolgy and Steps

3 Results

Task

TAGGING PODCASTS

Why do we want to tag podcasts?

- better categorize by main topics
- help listeners hone into specific episodes or podcasts
- draw audiences not exposed to a podcast before

Podcasts used

- 4 episode from 'The Argument' by The New York Times

Methodology and Steps

FOR TAGGING PODCASTS

Remove **alphanumeric** characters
and **lowercase** everything...

michelle goldberg i'm michelle goldberg. michelle cottle i'm michelle cottle. jamelle bouie i'm jamelle bouie. frank bruni and i'm frank bruni and this is the argument. the democratic convention is finally here if you can even call it a convention. like all else about a presidential campaign conducted in the shadow of a pandemic it's nothing like what happened in years past. does that negate its usefulness? what can and should democrats accomplish through this socially distanced virtual extravaganza? [...]

Remove stopwords

michelle goldberg i'm **michelle goldberg**. **michelle cottle** i'm **michelle cottle**. **janelle bouie** i'm **janelle bouie**. **frank bruni** and i'm **frank bruni** and this is the **argument**. the **democratic convention** is **finally** here if you can **even call** it a **convention**. **like** all **else** about a **presidential campaign** **conducted** in the **shadow** of a **pandemic** it's **nothing like** what **happened** in **years past**. does that **negate** its **usefulness**? what can and should **democrats accomplish** through this **socially distanced** **virtual extravaganza**? [...]

Lemmatization and keeping Nouns

Grouping words with similar meaning

BEFORE

michelle
goldberg
argument
democratic
convention
finally
even
call

AFTER

argument
democratic
convention

call

What is left to do?

- Building the **topic model**
- generate **word clouds** per topic

Results

WORDCLOUDS VS. DESCRIPTIONS

What Biden must do

A special round-table episode on what's at stake during this week's Democratic convention. Plus: a requiem for the "nerd Coachella."

A word cloud on a black background with words in various colors (teal, yellow, green). The words are: convention, thing, way, election, people, time.

Election, Convention

A word cloud on a black background with words in various colors (purple, teal, green). The words are: future, opponent, com, coalition, opinion, week.

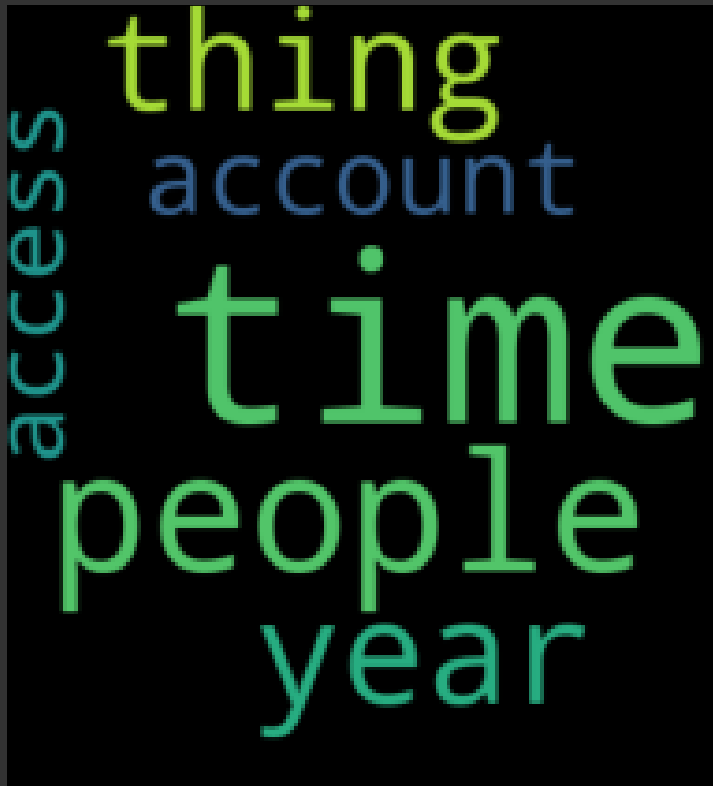
Coalition, Future

A word cloud on a black background with words in various colors (teal, yellow, blue, green). The words are: conversation, credibility, treatment, heart, force, range.

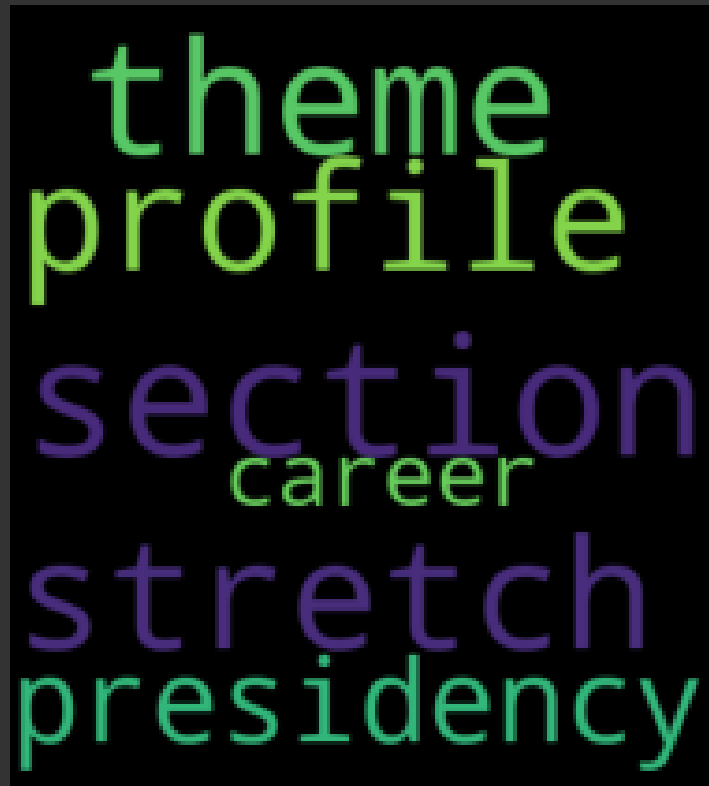
Treatment, Range

Can Republicans sell a whole new Trump?

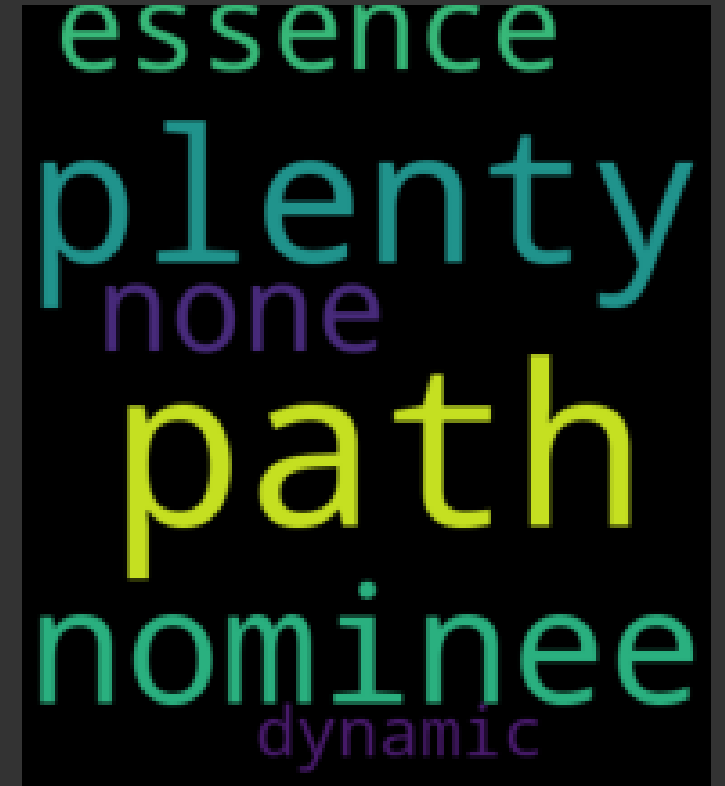
The Bulwark's Charlie Sykes joins the podcast to debate whether or not the convention will bring voters to Trump.



Time, People



Presidency, Section



Nominee, Path

Is Individualism America's Religion?

waytime
people
life
church
thing

Church, People

result
state
sense
mission
police
election

State, Police

recognition
mood
existence
surface
reflection
spokesman

Existence, Reflection

The case for a One-State solution

Peter Beinart joins the podcast to debate the future of Israelis and Palestinians.

people
state
thing
way
time lot

State, People

establishment
convention
concern
use
enthusiasm
citizen

Establishment, Citizen

columnist
stock
approval
case
institution
miscalculation

Case, Institution

Conclusion

Pros

- **Word clouds** resemble summary
- **Tag words** help categorization

Improvements

- **Remove** more words (eg. people)
- add **adjectives** for sentiment