Use of Natural Language Processing and Data Visualization for Ethics Pedagogy and Research

MICHEL MITRI

PROFESSOR EMERITUS

JAMES MADISON UNIVERSITY

Social/Technical Issues with Ethics in AI

Generative AI like GPT models can arguably pass the Turing Test

- ➔ they can fool us into thinking they are human

Dangers and limitations of AI

- Misinformation, privacy, employment, distributive justice, political propaganda, war, etc.
- In education – ease of cheating using ChatGPT
- Machine learning is only as good as the input data. ChatGPT is trained on the entire internet. GIGO
- Machine learning makes mistakes, even if the data is good

Intelligence is not the same as wisdom

Issues with Ethics in AI

Can AI ever pass the Winograd Schema Test?

- Common sense? Wisdom? Getting the facts right?
- See Hector Levesque: Common Sense, the Turing Test, and the Quest for Real AI
- Pronoun resolution: The trophy couldn't fit in the briefcase because it was too big.
  - Is "it" the briefcase or the trophy?
- ➔ Knowledge is important. Machine learning (pattern recognition) can't do it all. GOFAI

Can expert knowledge plus machine learning facilitate AI wisdom? AW?

https://storyanalyzer.org/

About Story Analyzer (SA)

Story Analyzer depends on two principal technologies

- Natural Language Processing – Stanford's CoreNLP
- Data Visualization – d3 (Data Driven Documents) and Google visualizations and maps

Using NLP, SA performs information extraction. Given a body of text, SA extracts the following:

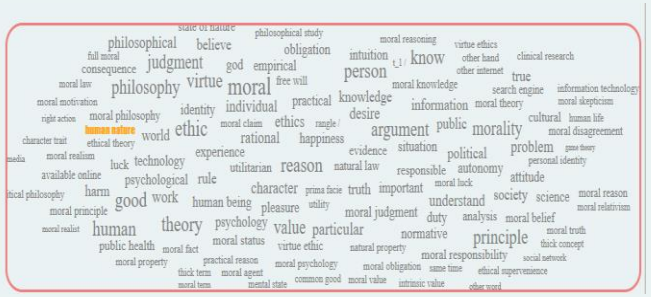
- People, organizations, and groups
- Places and Times
- Interactions between entities
- Linguistic relationships between concepts

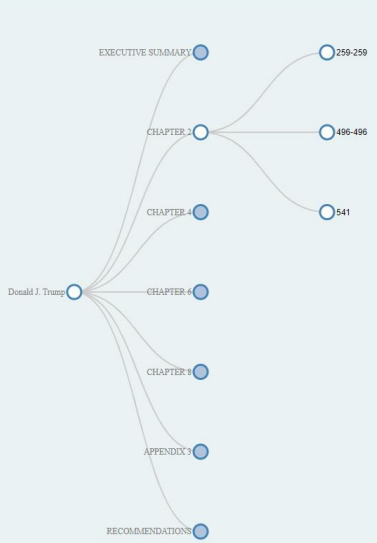
Stanford's CoreNLP

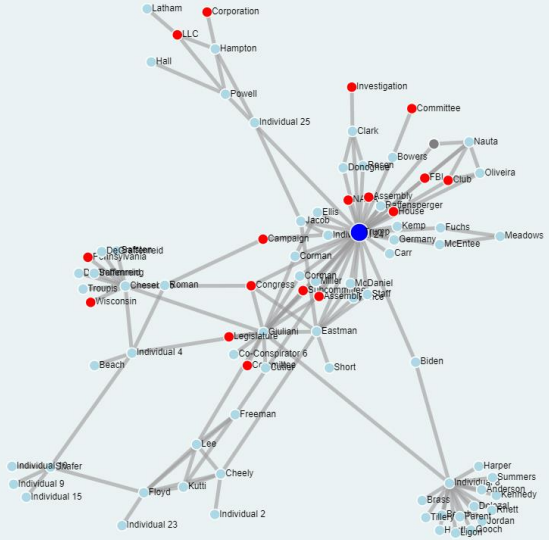
An open-source set of services that can do several things:

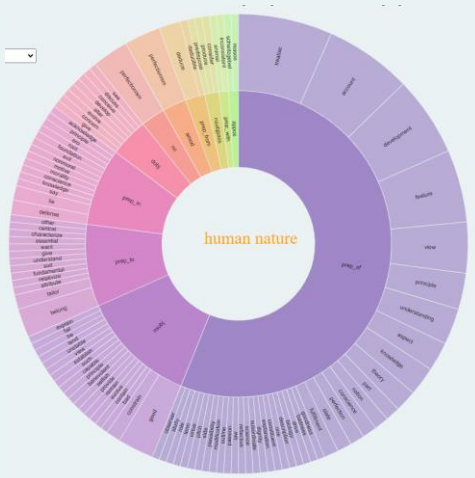
- Sentence splitting – breaking a text document into individual sentences
- Tokenizing a sentence (breaking it into individual "words")
- Identifying parts of speech (POS) within a sentence (nouns, verbs, adjectives, adverbs, etc.)
- Named entity recognition: Recognizing names of people, places, organizations
- Constituency parsing
- Dependency parsing
- Co-reference resolution – finding all expressions that refer to the same entity in a text. (e.g., finding connections between nouns and their associated pronouns)
- Temporal tagging – recognizing and normalizing temporal expressions

d3 Visualization types for SA

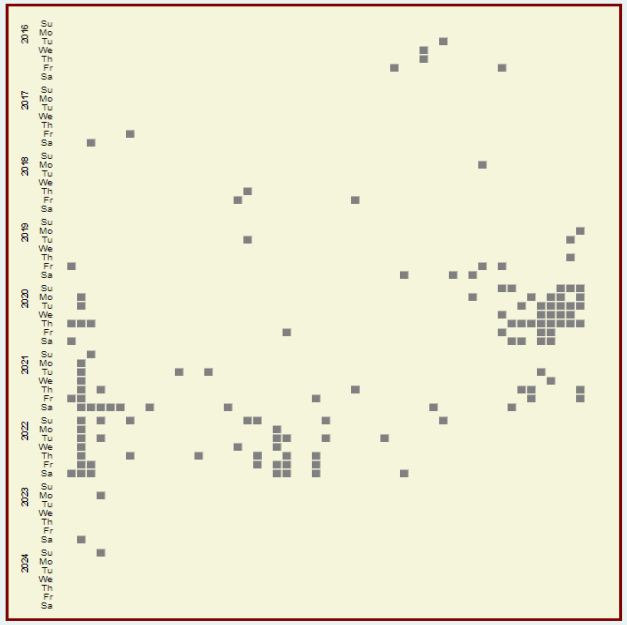




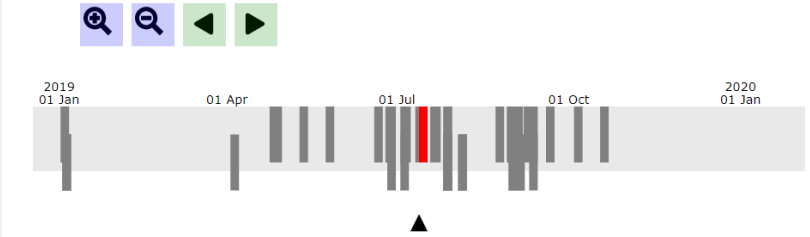




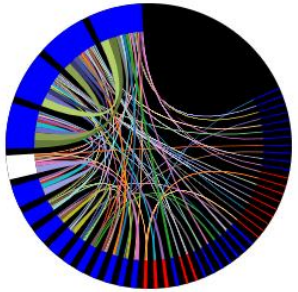
Calendar of Events



Timeline

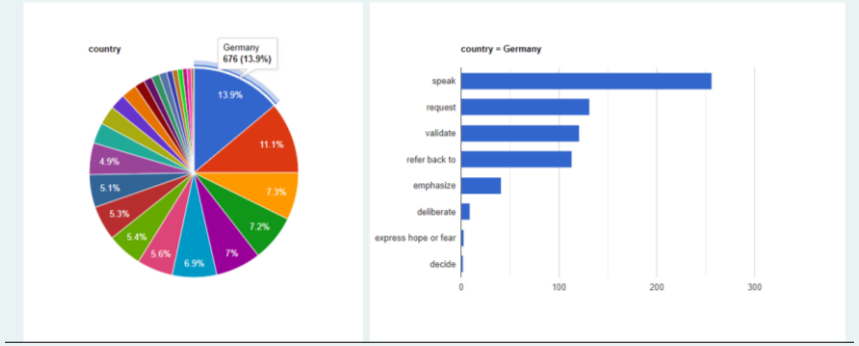


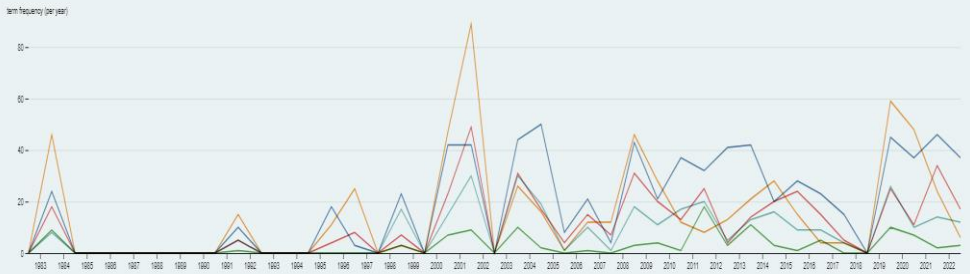
Interactions

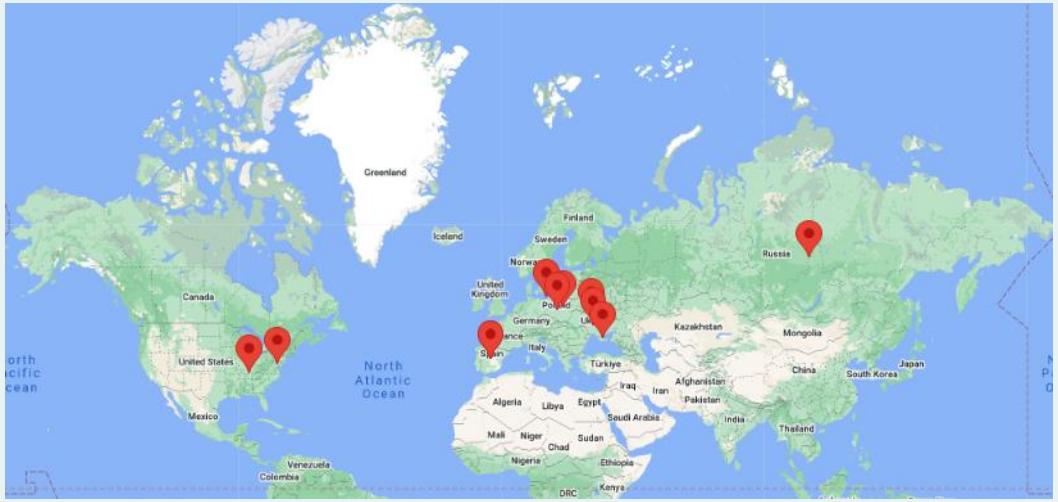


Google Visualization types for SA

People	#Mentions
Donald J. Trump	415
Rudolph Giuliani	97
Walt Nauta	89
John Eastman	61
Carlos De Oliveira	50
Kenneth Cheseboro	48
Harrison William Prescott Floyd	40
Robert David Cheely	40
David James Shafer	36
Trevian C. Kutti	34
Alston Latham	32
Joseph Biden	31
Sydney Powell	31
Jeffrey Clark	31
Cathleen Alston	

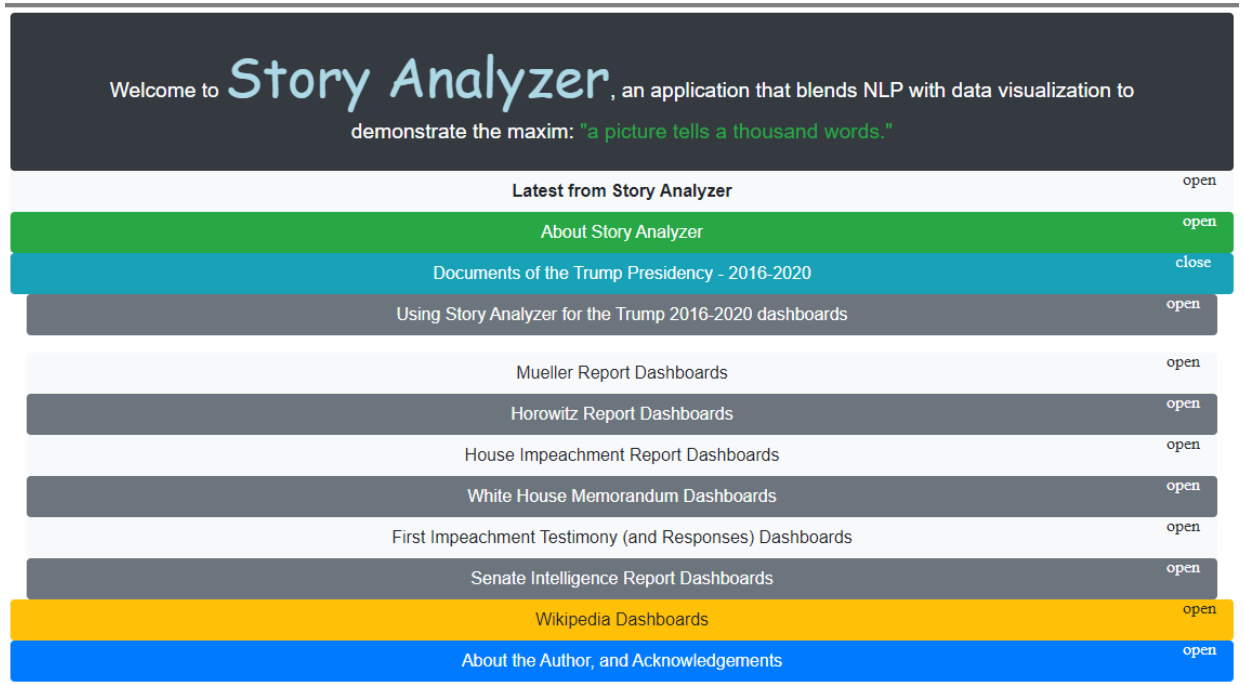






Early Story Analyzer Dashboards

- SA project began in 2017
- First robust dashboards:
  - Mueller Report
  - Horowitz Report
  - House Impeachment Report (2019)
  - Senate Intelligence Report
  - Various news and Wikipedia stories



Screenshot from https://storyanalyzer.org/

Example: House Impeachment Report

Story Analyzer - House Impeachment Report - Exec Summary Section 1 (pp. 12-27)

see https://intelligence.house.gov/uploadedfiles/the_trump-ukraine_impeachment_inquiry_report.pdf

Narrative and Highlighed Information

Narrative

EXECUTIVE SUMMARY

The impeachment inquiry into **Donald J. Trump**, the 45th President of the United States, uncovered a months-long effort by President **Trump** to use the powers of his office to solicit foreign interference on his behalf in the 2020 election. As described in this executive summary and the report that follows, President **Trump** 's scheme subverted U.S. foreign policy toward Ukraine and undermined our national security in favor of two politically motivated investigations that would help his presidential reelection campaign. The President demanded that the newly elected Ukrainian president, Volodymyr **Zelensky**, publicly announce investigations into a political rival that he apparently feared the most, former Vice President Joe Biden, and into a discredited theory that it was Ukraine, not Russia, that interfered in the 2016 presidential election. To compel the Ukrainian President to do his political bidding, President **Trump** conditioned two official acts on the public announcement of the investigations: a coveted White House visit and critical U.S. military assistance Ukraine needed to fight its Russian adversary.

During a July 25, 2019, call between President **Trump** and President **Zelensky**, President **Zelensky** expressed gratitude for U.S. military assistance. **President Trump immediately** responded **by** asking **President Zelensky to"** do **us a** favor **though"** and **openly** pressed **for** Ukraine **to** investigate **former Vice President** Biden **and the** 2016 **conspiracy** theory. **In** turn, **President Zelensky** assured **President Trump that he would** pursue **the** investigation **and** reiterated **his** interest **in the** White House meeting. Although President **Trump** 's scheme intentionally bypassed many career personnel, it was undertaken with the knowledge and approval of senior Administration officials, including the President 's Acting Chief of Staff Mick Mulvaney, Secretary of State Mike Pompeo, and Secretary of Energy Rick Perry. In fact, at a press conference weeks after public revelations about the

Highlighted Information

Interactions between President Zelensky-14-6 and Donald J. Trump-7-2

Sentence# 7) President **Trump** immediately responded by asking President **Zelensky** to" do us a favor though" and openly pressed for Ukraine to investigate former Vice President Biden and the 2016 conspiracy theory.

Sentence# 8) In turn, President **Zelensky** assured President **Trump** that he would pursue the investigation and reiterated his interest in the White House meeting.

Sentence# 42)

President **Trump** then asked President **Zelensky**" to look into" former Vice President Biden 's role in encouraging Ukraine to remove a prosecutor widely viewed by the United States and numerous European partners to be corrupt.

Sentence# 87) Days later, President **Trump** rescinded the plans for Vice President Pence to attend President **Zelensky** 's inauguration, which had not yet been scheduled

People, Groups, Interactions, and Narrative Web

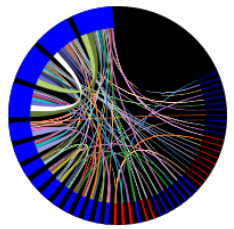
People

Person	Impact
Donald J. Trump-7-2	209
Ambassador Sondland-2-30	156
President Zelensky-14-6	124
President Trump-23-2	108
Rudy Giuliani-29-44	79
Ambassador Kurt Volker-11-90	37
Ambassador Taylor-7-100	36

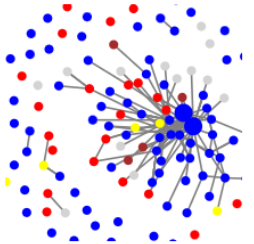
Groups

Group	Impact
NSC-21-25	13
Congress-21-57	13
Office of the President-20-11	11
Burisma-33-32	7
CNN-17-230	7
OMB-16-107	7
Congressional Intelligence Committees-24-257	6

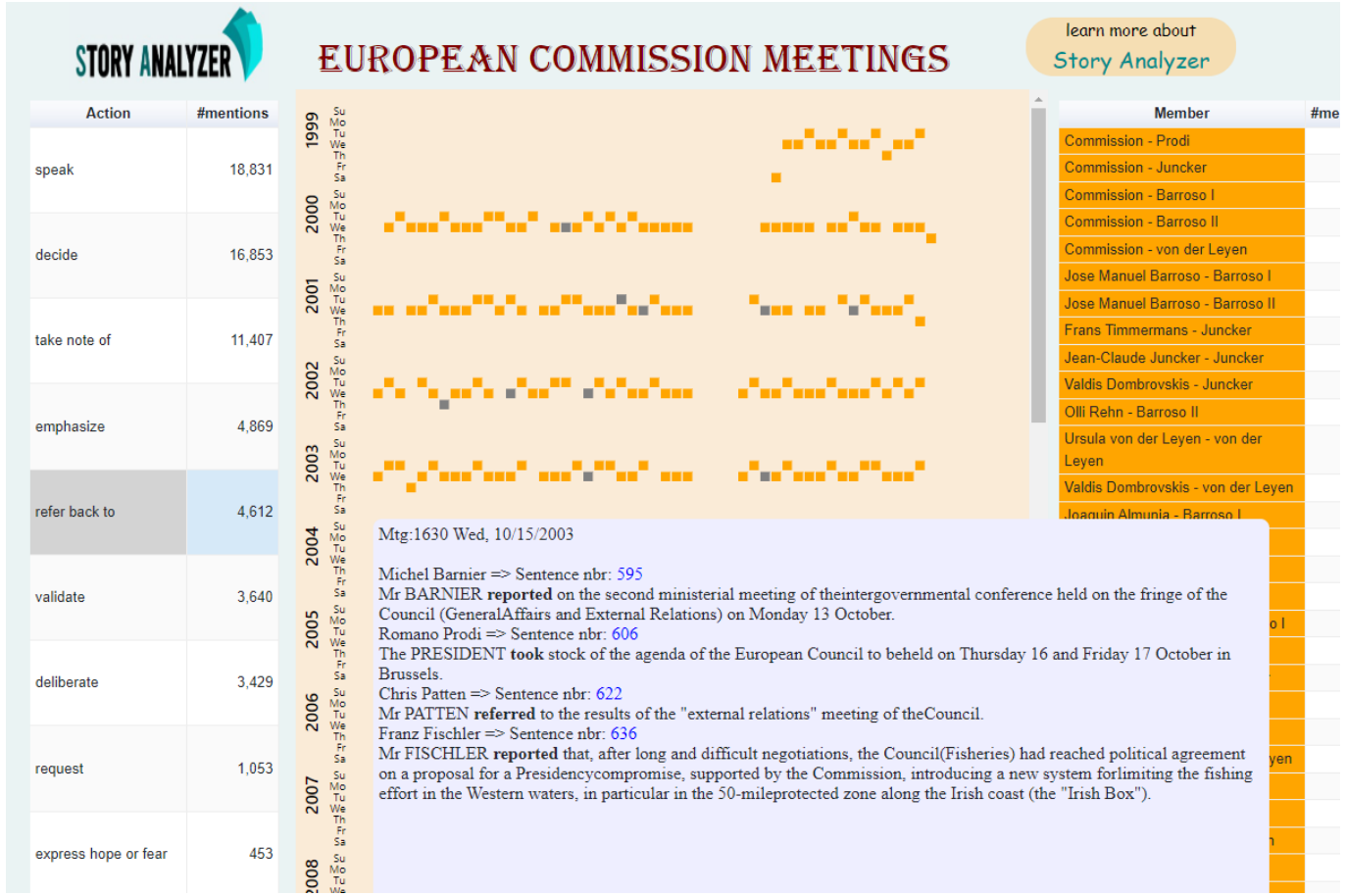
Interactions



Narrative Web



SA and Social Science Research



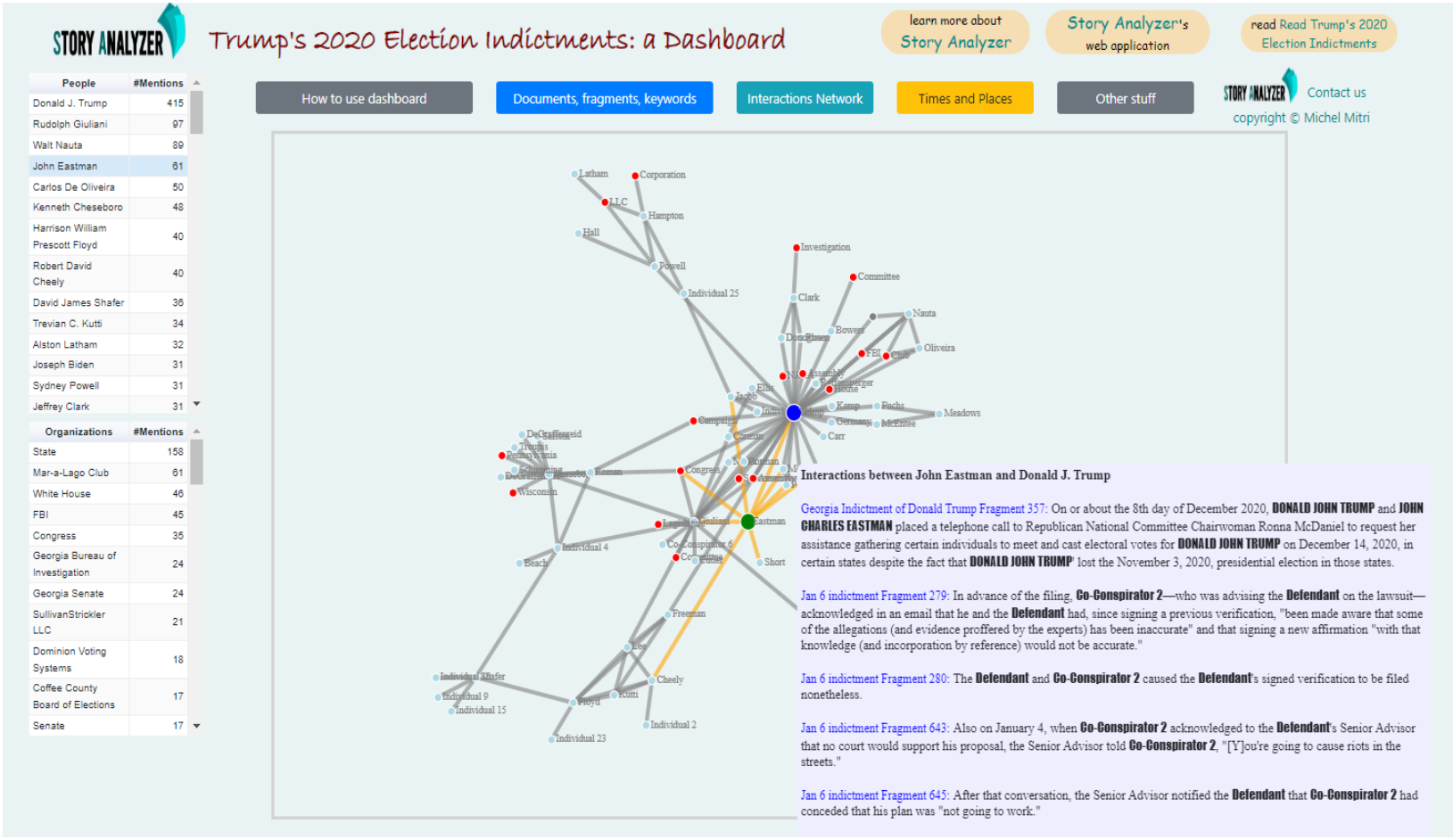
This dashboard supports the work of Dr. John Scherpereel, professor of political science at JMU.

Dr. Scherpereel is analyzing the minutes of 975 European Commission meetings going back to 1999 (almost 7 million words).

The dashboard depicts nine action categories and shows each sentence for that action either via the calendar or the list of EC member names.

https://storyanalyzer.org/ecmeetings/

January 6 and Trump Indictments



Jan 6 report – 840 pages, 360K words

https://storyanalyzer.org/jan6/Jan6Report.html

Trump indictments – Jan 6, Mar-a-Lago, Georgia

https://storyanalyzer.org/trumpindictments/

SA's Application to Student Ethics Project

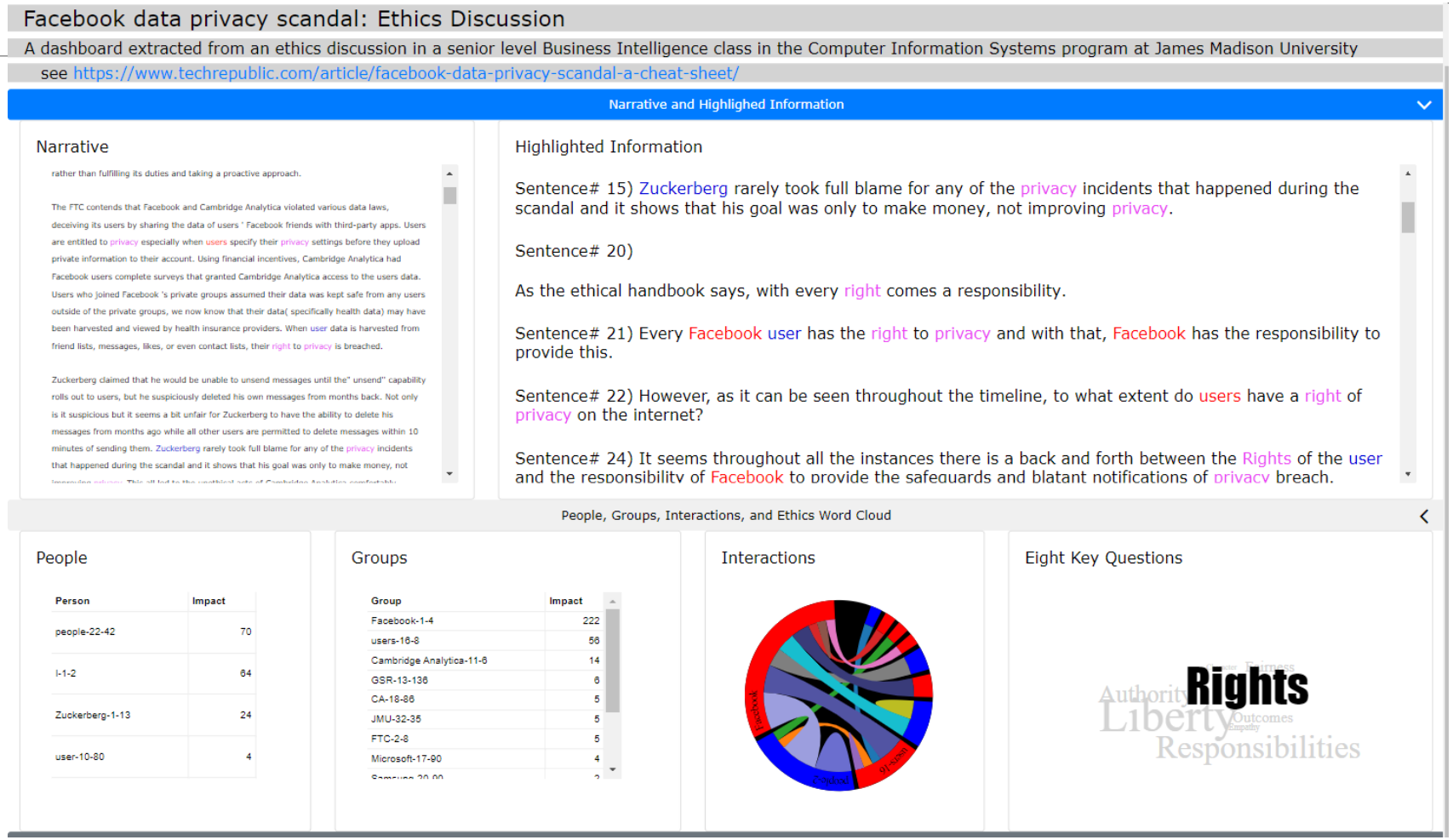
https://storyanalyzer.org/sawebsite/facebook%20ethics%20discussion%20spring%202020.html

JMU's Ethical Reasoning in Action (ERIA)

8 key questions: fairness, outcomes, responsibilities, character, liberty, empathy, authority, rights

Student project: evaluate Facebook & Cambridge Analytica scandal using 8KQ. Online discussion.

Dashboard of results:



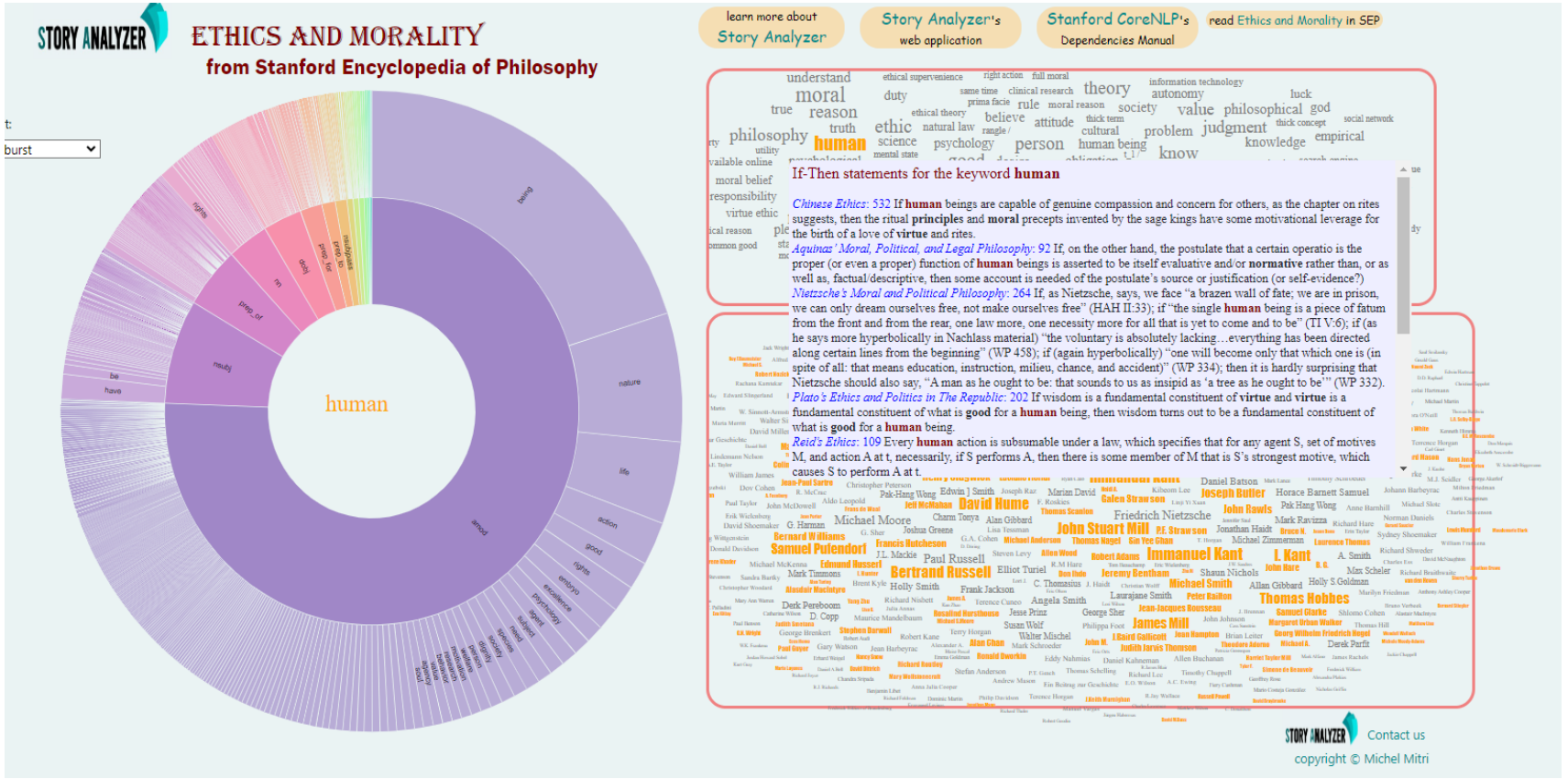
SA and Stanford Encyclopedia of Philosophy (SEP)

Extract of entire SEP corpus

Total 24 million words, 1800 articles

Currently four dashboards:

1. Truth – 24 articles
2. Justice – 31 articles
3. Political Theory – 52 articles
4. Ethics – 94 articles (1.5 million words)

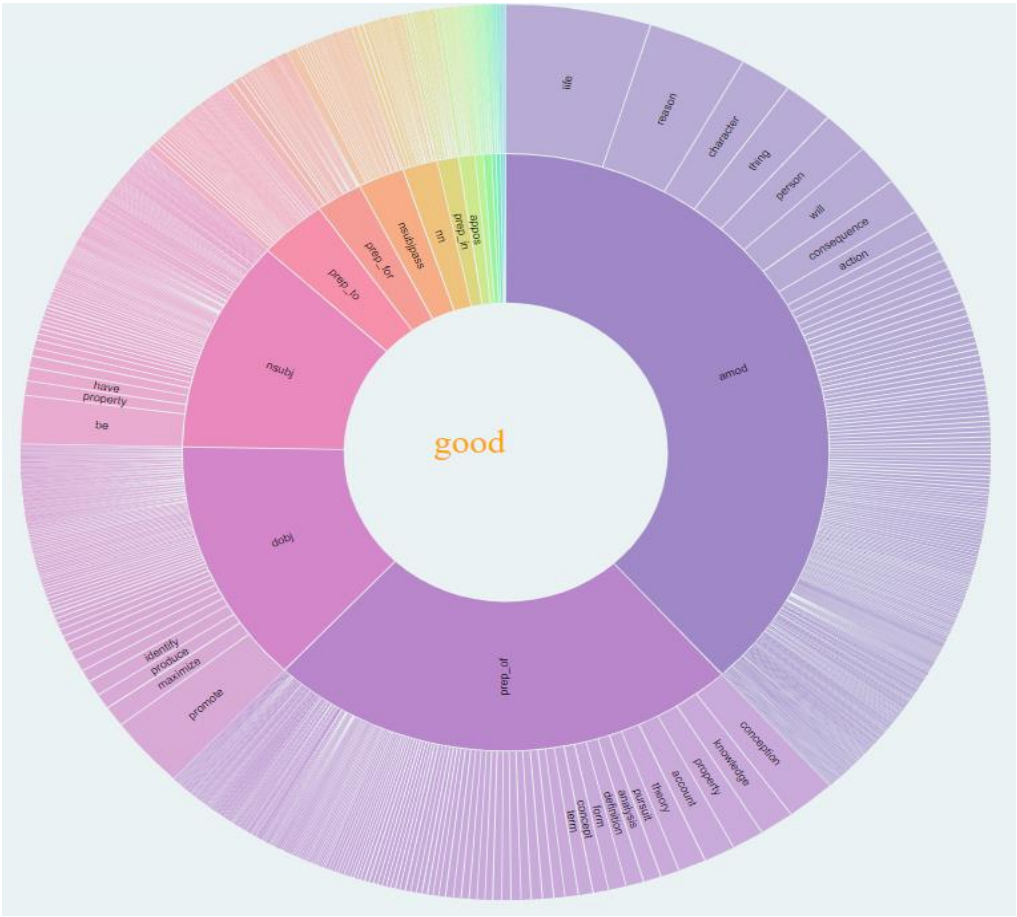


Using SEP Dashboards

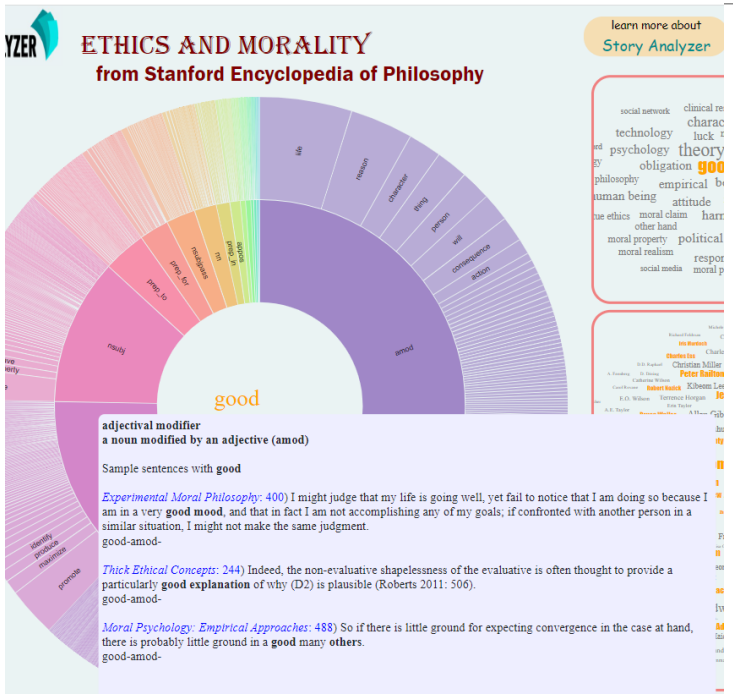
As you hover over keywords in the top word cloud, all sentences of "if-then" structure with that keyword are shown.

Clicking a keyword displays a sunburst visualization.

This shows linguistic relations (via CoreNLP's dependency parsing) of terms rooted at the keywor.



Using SEP Dashboards



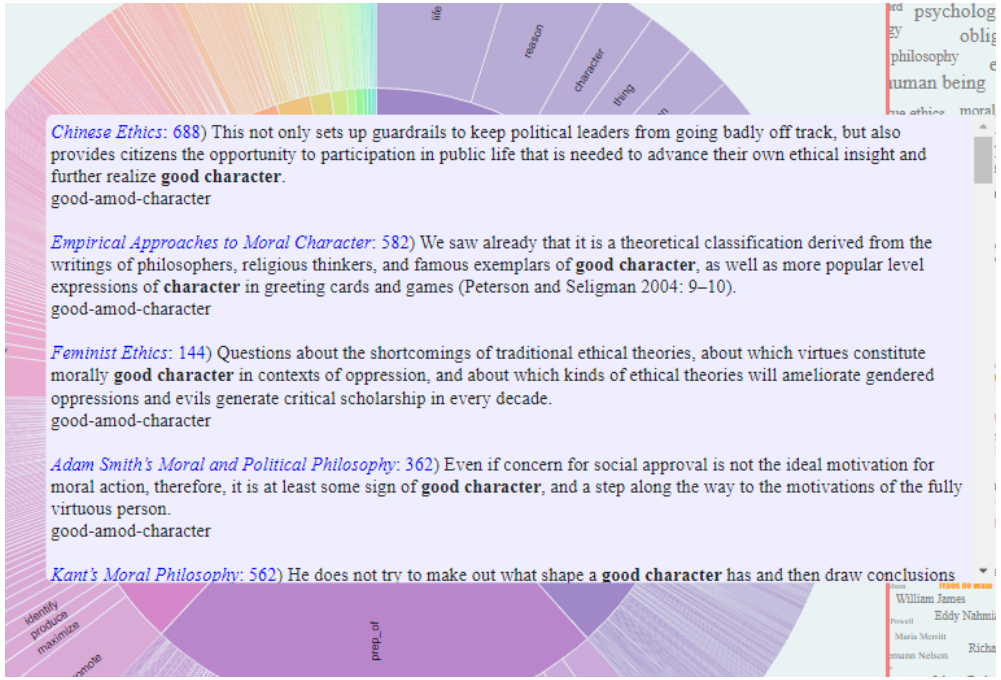
The **sunburst** center is the selected word. The middle ring shows dependency relationship types, and the outer ring shows word connected to the keyword via these relationships.

Some CoreNLP dependency relationships:

- **amod** – adjectival modifier
- **prep** (of, in, from, for, etc.) – prepositions
- **nn** – compound noun
- **nsubj** – nominal subject
- **dobj** – direct object
- Lots of others

If you hover over an item in the middle ring, you'll see an explanation of the dependency type and some sample sentences involving that type.

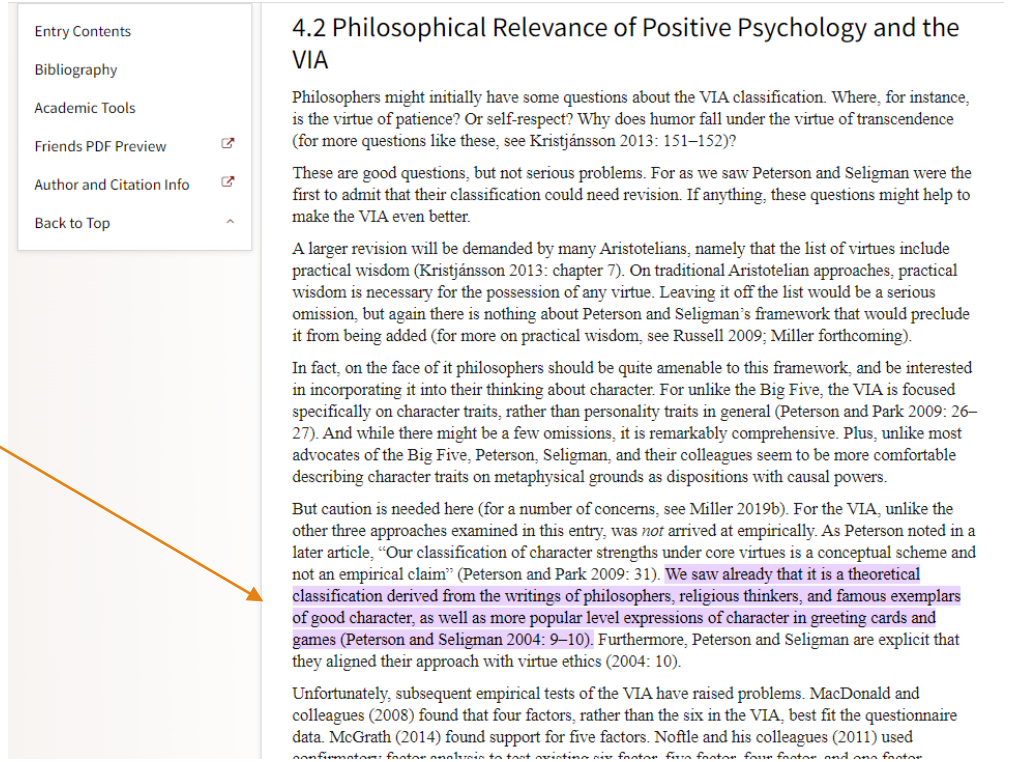
Using SEP Dashboards



Clicking a segment in the outer ring (**character** in this case) will show all sentences with the specified dependency relationship between the keyword and the outer-ring word.

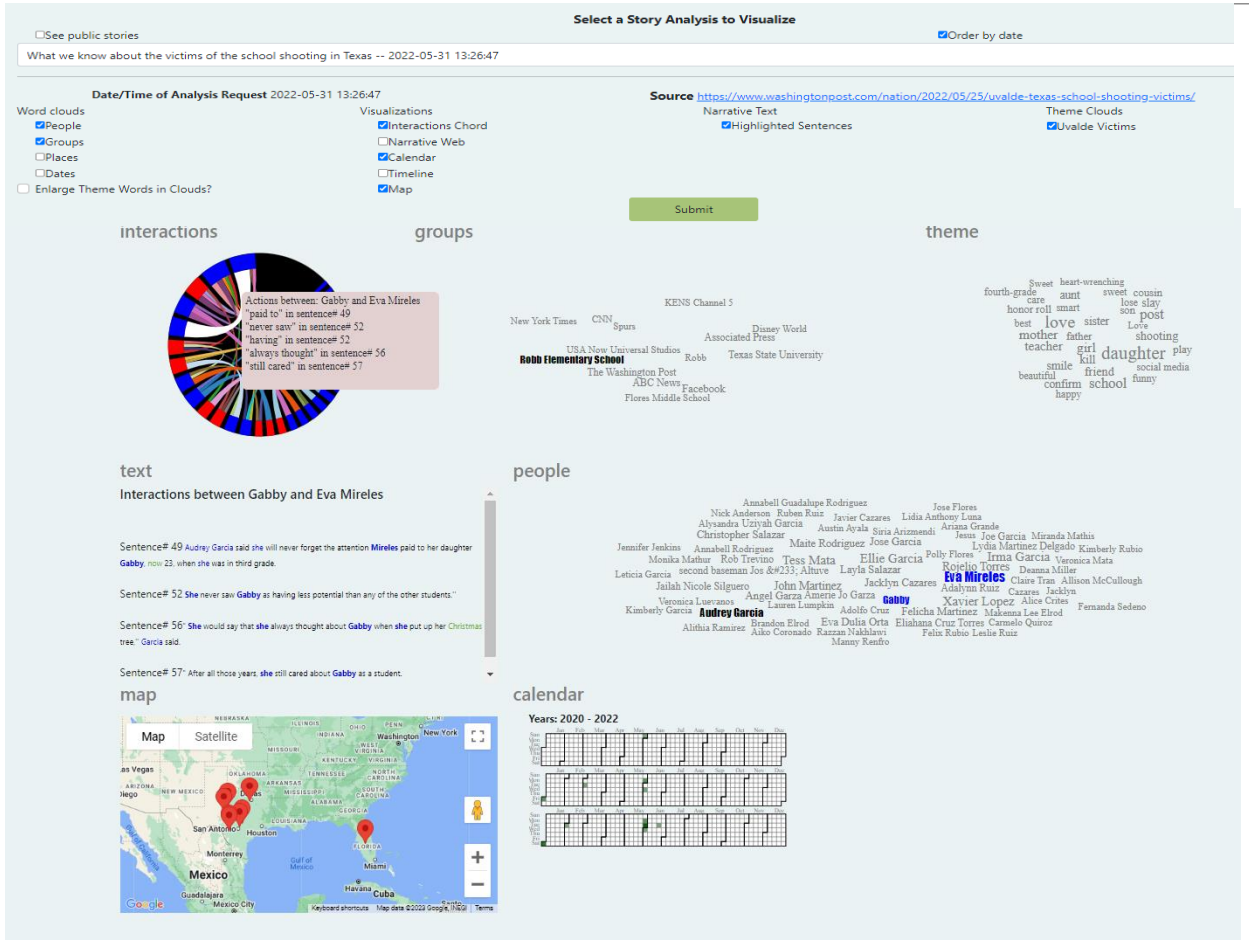
For a sentence number of an article, if you hover over the blue italicized text, you'll see expanded text showing surrounding sentences. Click the blue text to freeze the expansion in place.

Using SEP Dashboards



Finally, clicking the "see in browser" link of the expansion text should take you to the specific SEP article and the specific sentence. This doesn't work with all browsers.

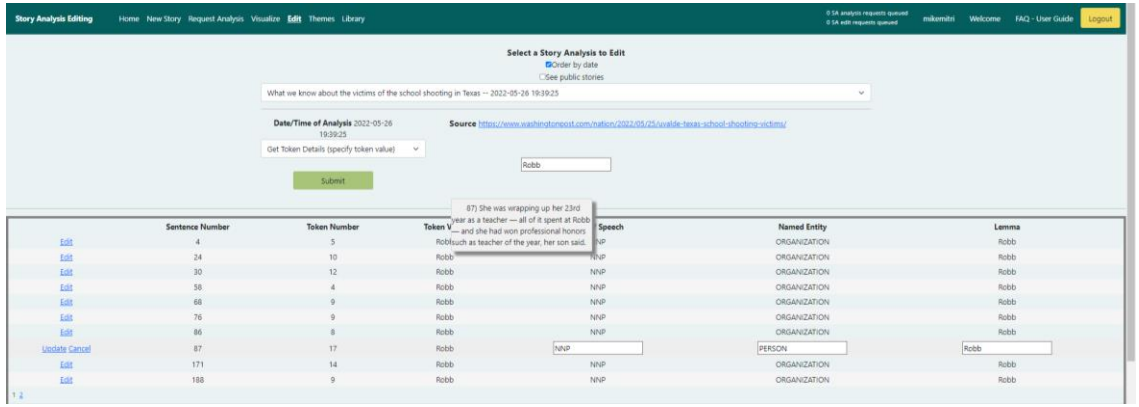
Want to build your own SA dashboards?



https://app.storyanalyzer.org

Display interactive visualizations in dashboard

Edit/Correct NLP errors



Limitations, differences, possibilities, approaches

Machine learning's Accuracy limitations

- POS – 95%+
- NER, Dependency parsing – 80%+
- Coreference resolution – 60%+
- ChatGPT – lots of demonstrable errors
- Even OCR has errors

My software has bugs. It's a work in progress.

Purpose of ChatGPT vs. SA & CoreNLP – generative (writing, creating) vs analytical (reading, understanding)

Inductive reasoning vs. knowledge-based deduction (ML vs "expert systems")

Reprise – Common sense? Wisdom? Getting it right? GIGO?

Future Goals for SA

Can SA help to evaluate the truth of falseness of a story? Can SA become a "fact-checker"?

Philosophical theories of truth:

- Correspondence theory
  - A statement is true if it corresponds with the facts
  - Compare stories to research findings (e.g., Our World in Data), and Google FactCheck (from Politico)
- Coherence theory
  - A statement is true if it coheres with other well-accepted truths
  - Compare stories against each other
- Pragmatic theory
  - A statement is true if it "works", if it produces beneficial results
  - Prescriptive analytics – decision theory, utility measures, etc.

Questions?