

TEXT SUMMARIZATION OF CONGRESSIONAL BILLS

Connor Cabrey

Dalia Habiby

AGENDA

Introduction

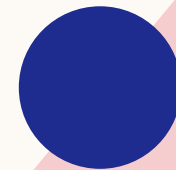
Data

Pre-Processing

NLP Approach

Results

Conclusion / Discussion



INTRODUCTION

Using data from the 117th Congress Second Session, we will apply text summarization methods in order to effectively summarize bills that are incredibly lengthy. Applying extractive text summarization methods will hopefully begin to solve this problem.

DATA

- Congressional Library
- 117th Congress 2nd Session
- 4674 Individual Bills
 - XML format
 - Extracted text elements
 - Many iterations of the same bill
- 152 “enrolled” bills

GovInfo

[Browse](#)[About](#)[Developers](#)[Features](#)[Help](#)[Feedback](#)

Q SEARCH ▾

Bulk Data Repository

[f](#)[t](#)[e](#)[r](#)

HELP

GPO makes select collections available in bulk in a machine-readable format (i.e. XML) via our Bulk Data Repository. The top level directory for select collections, such as the Federal Register and Code of Federal Regulations, also includes a Resources directory that contains the XML schema, XSL stylesheet, and user guide.

For more information and to report any issues, please see our [GitHub repository](#).

Bulk Data

Congressional Bills

117th Congress (2021-2022)

2nd Session

hr

Name	Last Modified (GMT)	Size
BILLS-117hr1036rs.xml	29-Dec-2022 00:47	7.4 kB
BILLS-117hr1057eh.xml	29-Dec-2022 00:47	14.7 kB
BILLS-117hr1057enr.xml	29-Dec-2022 00:47	14.8 kB
BILLS-117hr1057rds.xml	29-Dec-2022 00:47	14.7 kB
BILLS-117hr1066eh.xml	28-Dec-2022 22:47	4.1 kB

CLEANING / PRE-PROCESSING

- Required 'ElementTree' Python Package
- XML is inherently hierarchical and is best represented as a tree
- Different levels are represented with different tags
- Individual nodes on the tree are called elements

```
import xml.etree.ElementTree as ET

# Load the XML file
tree = ET.parse("../xml_files\\BILLS-117hr8pcs.xml")

# Get the root element
root = tree.getroot()

# Collect all unique root tags
roots = set()
for elem in root.iter():
    roots.add(elem.tag)

# Print the unique root tags
print("Possible roots in the XML file:")
for root_tag in roots:
    print(root_tag)
```

Possible roots in the XML file:

enum	enum
metadata	subclause
action-desc	date
subsection	action-date
bill	quote
external-xref	current-chamber
text	rights
legis-type	congress
header	title
official-title	format
quoted-block	role
paragraph	attestation
subparagraph	legis-num
form	action
attestation-group	enum-in-header
dublinCore	attestation-date
attestor	clause
calendar	legis-body
section	endorsement
session	language
publisher	short-title
after-quoted-block	distribution-code

LEXRANK SUMMARIZER

- Unsupervised, graph-based algorithm
- Inspired by Google's PageRank
- Based on eigenvector centrality
- Uses a connectivity matrix based on intra-sentence cosine similarity

	sentence 1	sentence 2	sentence 3
sentence 1	0	1	0
sentence 2	1	0	0
sentence 3	1	1	0

example connectivity matrix

1st iteration

	sentence 1	sentence 2	sentence 3			standardize
sentence 1	0	1	0	1	1	0,408
sentence 2	1	0	0	1	1	0,408
sentence 3	1	1	0	1	2	0,816

normalized value 2,44949

2nd iteration

	sentence 1	sentence 2	sentence 3			standardize
sentence 1	0	1	0	0,408	0,408	0,167
sentence 2	1	0	0	0,408	0,408	0,167
sentence 3	1	1	0	0,816	1,633	0,667

normalized value 1,732

SUMMARY EXAMPLE 1

Summary

'Congress finds the following:Richard A. Pittman was born on May 26, 1945, in French Camp, California.Richard A. Pittman enlisted in the Marine Corps in 1965 and was later promoted to Lance Corporal.On July 24, 1966, Lance Corporal Pittman's unit, First Platoon, Company I, Third Battalion, Fifth Marines, was moving along a trail near the Demilitarized Zone in the Republic of Vietnam when the Marines ahead of them came under attack.Upon hearing the attack, Lance Corporal Pittman took a machine gun and several belts of ammunition and rushed to aid his comrades. As he advanced, Lance Corporal Pittman encountered intense small-arms fire at point-blank range. He was able to return fire and silence the enemy combatants. Continuing towards the attacked Marines, Lance Corporal Pittman came under heavy fire from two automatic weapons, which he was able to destroy.When Lance Corporal Pittman learned there were additional wounded Marines further along the trail, he endured enemy mortar and gunfire to get to them. As he reached them, he faced an attack by 30 to 40 enemy fighters.Lance Corporal Pittman used a combination of machine-gun fire, an enemy submachine gun, a pistol from a fallen comrade, and a grenade to cause enemy forces to withdraw.'

Original Bill

'Congress finds the following:Richard A. Pittman was born on May 26, 1945, in French Camp, California.Richard A. Pittman enlisted in the Marine Corps in 1965 and was later promoted to Lance Corporal.On July 24, 1966, Lance Corporal Pittman's unit, First Platoon, Company I, Third Battalion, Fifth Marines, was moving along a trail near the Demilitarized Zone in the Republic of Vietnam when the Marines ahead of them came under attack.Upon hearing the attack, Lance Corporal Pittman took a machine gun and several belts of ammunition and rushed to aid his comrades. As he advanced, Lance Corporal Pittman encountered intense small-arms fire at point-blank range. He was able to return fire and silence the enemy combatants. Continuing towards the attacked Marines, Lance Corporal Pittman came under heavy fire from two automatic weapons, which he was able to destroy.When Lance Corporal Pittman learned there were additional wounded Marines further along the trail, he endured enemy mortar and gunfire to get to them. As he reached them, he faced an attack by 30 to 40 enemy fighters.Lance Corporal Pittman used a combination of machine-gun fire, an enemy submachine gun, a pistol from a fallen comrade, and a grenade to cause enemy forces to withdraw. Due to his courageous actions, many of his wounded comrades were saved.In recognition of his bravery that day, Richard A. Pittman, who was then a Sergeant, was awarded the Medal of Honor by President Lyndon B. Johnson on May 14, 1968.Richard A. Pittman retired in 1988 as a Master Sergeant, after having served in the Marine Corps for more than two decades.The Department of Veterans Affairs community-based outpatient clinic in French Camp, California, shall after the date of the enactment of this Act be known and designated as the Any reference in a law, map, regulation, document, paper, or other record of the United States to the clinic referred to in subsection (a) shall be deemed to be a reference to the '

SUMMARY EXAMPLE 2

Summary

'This Act may be known as the All products of the Russian Federation classified under chapter 27 of the Harmonized Tariff Schedule of the United States shall be banned from importation into the United States, in a manner consistent with any implementation actions issued under Executive Order 14066 (87 Fed. A motion to reconsider the vote by which the motion is disposed of shall not be in order. All points of order against the joint resolution and against its consideration are waived. A joint resolution of disapproval introduced in the Senate shall be referred to the Committee on Finance.If the Committee on Finance has not reported such joint resolution of disapproval within 10 days on which the Senate is in session after the date of referral of such joint resolution, that committee shall be discharged from further consideration of such joint resolution and the joint resolution shall be placed on the appropriate calendar.Notwithstanding Rule XXII of the Standing Rules of the Senate, it is in order at any time after the Committee on Finance reports the joint resolution of disapproval to the Senate or has been discharged from its consideration (even though a previous motion to the same effect has been disagreed to) to move to proceed to the consideration of the joint resolution, and all points of order against the joint resolution (and against consideration of the joint resolution) shall be waived. An amendment to, or a motion to postpone, or a motion to proceed to the consideration of other business, or a motion to recommit the joint resolution of disapproval is not in order.The vote on passage shall occur immediately following the conclusion of the debate on the joint resolution of disapproval and a single quorum call at the conclusion of the debate, if requested in accordance with the rules of the Senate.Appeals from the decisions of the Chair relating to the application of the rules of the Senate, as the case may be, to the procedure relating to the joint resolution of disapproval shall be decided without debate.Debate in the Senate of any veto message with respect to the joint resolution of disapproval, including all debatable motions and appeals in connection with such joint resolution, shall be limited to 10 hours, to be equally divided between, and controlled by, the majority leader and the minority leader or their designees.Except as otherwise provided in this subsection, the following procedures shall apply in the Senate to a joint resolution of disapproval:Except as provided in subparagraph (B), a joint resolution of disapproval that has passed the House of Representatives shall, when received in the Senate, be referred to the Committee on Finance for consideration in accordance with this subsection.If a joint resolution of disapproval was introduced in the Senate before receipt of a joint resolution of disapproval that has passed the House of Representatives, the joint resolution from the House of Representatives shall, when received in the Senate, be placed on the calendar.'

This Act may be known as the All products of the Russian Federation classified under chapter 27 of the Harmonized Tariff Schedule of the United States shall be banned from importation into the United States, in a manner consistent with any implementation actions issued under Executive Order 14066 (87 Fed. Reg. 13625; relating to prohibiting certain imports and new investments with respect to continued Russian Federation efforts to undermine the sovereignty and territorial integrity of Ukraine).The President is authorized to terminate the prohibition on importation of energy products of the Russian Federation under section 2 if the President submits to Congress a certification under subsection (c). Such termination shall take effect beginning on the date that is 90 calendar days after the date of submission of such certification, unless there is enacted into law during such 90-day period a joint resolution of disapproval.The President shall, not later than 45 calendar days before submitting a certification under subsection (a)—consult with—the Committee on Ways and Means and the Committee on Foreign Affairs of the House of Representatives; and the Committee on Finance and the Committee on Foreign Relations of the Senate; andsubmit to all such committees a report that explains the basis for the determination of the President contained in such certification.A certification under this subsection is a certification in writing that—indicates that the President proposes to terminate under subsection (a) the prohibition under section 2; andcontains a determination of the President that the Russian Federation—has reached an agreement to withdraw Russian forces and for the cessation of military hostilities that is accepted by the free and independent government of Ukraine;poses no immediate military threat of aggression to any North Atlantic Treaty Organization member; andrecognizes the right of the people of Ukraine to independently and freely choose their own government. For purposes of this section, the term that does not have a preamble;the title of which is as follows: the matter after the resolving clause of which is as follows: During a period of 5 legislative days beginning on the date that a certification under subsection (c) is submitted to Congress, a joint resolution of disapproval may be introduced in the House of Representatives by the majority leader or the minority leader.During a period of 5 days on which the Senate is in session beginning on the date that a certification under subsection (c) is submitted to Congress, a joint resolution of disapproval may be introduced in the Senate by the majority leader (or the majority leader's designee) or the minority leader (or the minority leader's designee). If a committee of the House to which a joint resolution of disapproval has been referred has not reported such joint resolution within 10 legislative days after the date of referral, that committee shall be discharged from further consideration thereof. Beginning on the third legislative day after each committee to which a joint resolution of disapproval has been referred reports it to the House or has been discharged from further consideration thereof, it shall be in order to move to proceed to consider the joint resolution in the House. All points of order against the motion are waived. Such a motion shall not be in order after the House has disposed of a motion to proceed on a joint resolution with regard to the same certification. The previous question shall be considered as ordered on the motion to its adoption without intervening motion. The motion shall not be debatable. A motion to reconsider the vote by which the motion is disposed of shall not be in order. The joint resolution shall be considered as read. All points of order against the joint resolution and against its consideration are waived. The previous question shall be considered as ordered on the joint resolution to final passage without intervening motion except two hours of debate equally divided and controlled by the sponsor of the joint resolution (or a designee) and an opponent. A motion to reconsider the vote on passage of the joint resolution shall not be in order. A joint resolution of disapproval introduced in the Senate shall be referred to the Committee on Finance.If the Committee on Finance has not reported such joint resolution of disapproval within 10 days on which the Senate is in session after the date of referral of such joint resolution, that committee shall be discharged from further consideration of such joint resolution and the joint resolution shall be placed on the appropriate calendar.Notwithstanding Rule XXII of the Standing Rules of the Senate, it is in order at any time after the Committee on Finance reports the joint resolution of disapproval to the Senate or has been discharged from its consideration (even though a previous motion to the same effect has been disagreed to) to move to proceed to the consideration of the joint resolution, and all points of order against the joint resolution (and against consideration of the joint resolution) shall be waived. The motion to proceed is not debatable. The motion is not subject to a motion to postpone. A motion to reconsider the vote by which the motion is agreed to or disagreed to shall not be in order. If a motion to proceed to the consideration of the joint resolution of disapproval is agreed to, the joint resolution shall remain the unfinished business until disposed of.Debate on the joint resolution of disapproval, and on all debatable motions and appeals in connection therewith, shall be limited to not more than 10 hours, which shall be divided equally between the majority and minority leaders or their designees. A motion to further limit debate is in order and not debatable. An amendment to, or a motion to postpone, or a motion to proceed to the consideration of other business, or a motion to recommit the joint resolution of disapproval is not in order.The vote on passage shall occur immediately following the conclusion of the debate on the joint resolution of disapproval and a single quorum call at the conclusion of the debate, if requested in accordance with the rules of the Senate.Appeals from the decisions of the Chair relating to the application of the rules of the Senate, as the case may be, to the procedure relating to the joint resolution of disapproval shall be decided without debate.Debate in the Senate of any veto message with respect to the joint resolution of disapproval, including all debatable motions and appeals in connection with such joint resolution, shall be limited to 10 hours, to be equally divided between, and controlled by, the majority leader and the minority leader or their designees.Except as otherwise provided in this subsection, the following procedures shall apply in the Senate to a joint resolution of disapproval:Except as provided in subparagraph (B), a joint resolution of disapproval that has passed the House of Representatives shall, when received in the Senate, be referred to the Committee on Finance for consideration in accordance with this subsection.If a joint resolution of disapproval was introduced in the Senate before receipt of a joint resolution of disapproval that has passed the House of Representatives, the joint resolution from the House of Representatives shall, when received in the Senate, be placed on the calendar. If this subparagraph applies, the procedures in the Senate with respect to a joint resolution of disapproval introduced in the Senate that contains the identical matter as the joint resolution of disapproval that passed the House of Representatives shall be the same as if no joint resolution of disapproval had been received from the House of Representatives, except that the vote on passage in the Senate shall be on the joint resolution of disapproval that passed the House of Representatives.This subsection is enacted by Congress—as an exercise of the rulemaking power of the Senate and the House of Representatives, respectively, and as such is deemed a part of the rules of each House, respectively, but applicable only with respect to the procedure to be followed in that House in the case of a joint resolution of disapproval, and supersedes other rules only to the extent that it is inconsistent with such rules; andwith full recognition of the constitutional right of either House to change the rules (so far as relating to the procedure of that House) at any time, in the same manner, and to the same extent as in the case of any other rule of that House.'

RESULTS

All 152 Bills

Figure 1: Readability Score Difference: All Bills

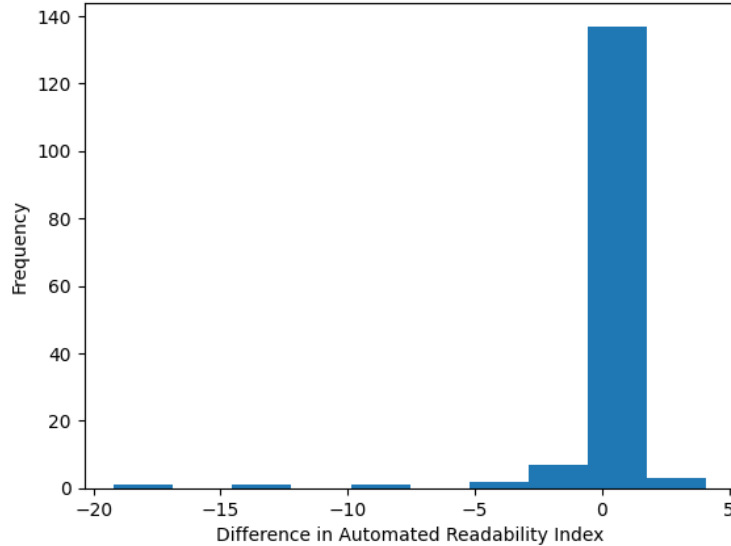
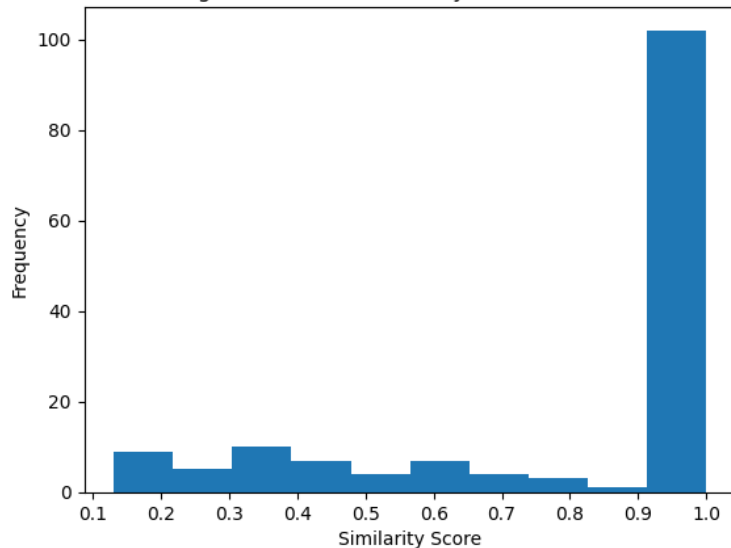


Figure 2: Cosine Similarity Scores: All Bills



- Overall poor performance as a method
- Readability was lower for summaries than original bills
- Summary lengths tended to be just as long as the original bills
- Different methods for determining 'sentences' within the model could lead to better performance.

22 Short Summaries

Figure 3: Readability Score Difference: Shorter Summaries

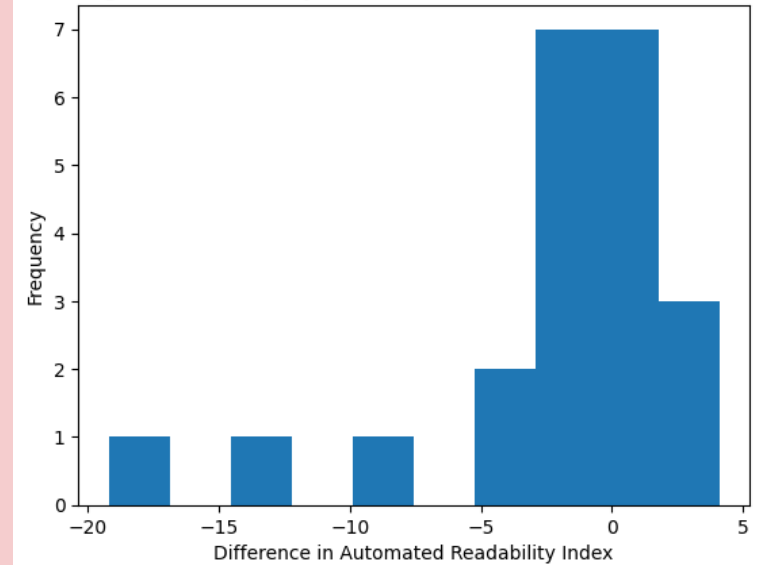
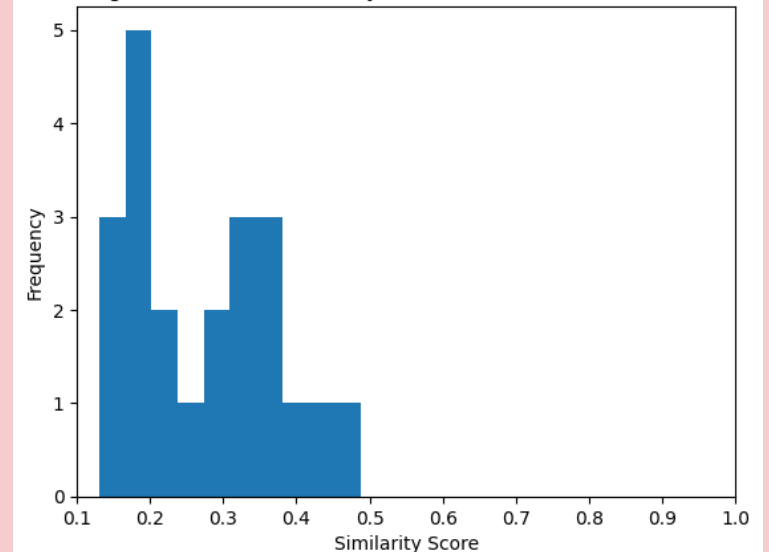


Figure 4: Cosine Similarity Scores For Shorter Summaries





DISCUSSION/ QUESTIONS

REFERENCES

- Bennani, H., & Neuenkirch, M. (2022). Too complex to digest? federal tax bills and their processing in us financial markets.
- Discover u.s. government information. (n.d.). <https://www.govinfo.gov/>
- Kanapala, A., Pal, S., & Pamula, R. (2017). Text summarization from legal documents: A survey. *Artificial Intelligence Review*.
- Key to versions of printed legislation. U.S. Senate: Key to Versions of Printed Legislation. (n.d.). <https://www.senate.gov/legislative/KeytoVersionsofPrintedLegislation.htm>
- Kornilova, A., & Eidelman, V. (2019). Billsum: A corpus for automatic summarization of us legislation. *arXiv preprint arXiv:1910.00523*.
- Md, Abdul Quadir, Raghav V. Anand, Senthilkumar Mohan, Christy Jackson Joshua, Sabhari S. Girish, Anthra Devarajan, and Celestine Iwendi. (2023). “Data-Driven Analysis of Privacy Policies Using LexRank and KL Summarizer for Environmental Sustainability” *Sustainability* 15, no. 7: 5941. <https://doi.org/10.3390/su15075941>
- Salima Lamsiyah, Abdelkader El Mahdaouy, Bernard Espinasse, Saïd El Alaoui Ouatik, An unsupervised method for extractive multi-document summarization based on centroid approach and sentence embeddings, *Expert Systems with Applications*, Volume 167, 2021, 114152, ISSN 0957-4174, <https://doi.org/10.1016/j.eswa.2020.114152>
- Zalwert, Maciej. “LexRank Algorithm Explained: A Step-by-Step Tutorial with Examples.” *Medium*, 10 Apr. 2022, maciejzalwert.medium.com/lexrank-algorithm-explained-a-step-by-step-tutorial-with-examples-3d3aa0297c57.