IDA SCIENCE AND TECHNOLOGY POLICY INSTITUTE

Text Explorer

William E. J. Doane, PhD

Assistant Director

wdoane@ida.org

10/29/2024

Science and Technology Policy Institute

1701 Pennsylvania Ave, NW ● Suite 500 ● Washington, DC 20006-5825

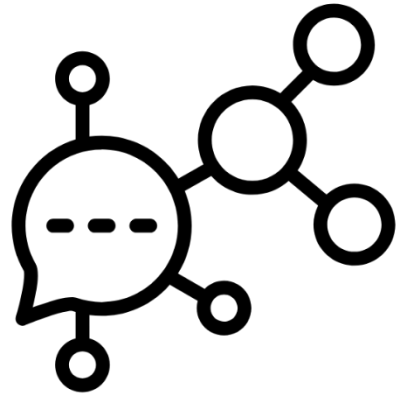
IDA's Science and Technology Policy Institute

- A federally funded research and development center
- Chartered by U.S. Congress
- Supports the White House Office of Science and Technology Policy (OSTP)
- Provides objective, rigorous analyses for U.S. Federal Government agencies related to science and technology
- Develops tools for internal use to support our analyses



Text Explorer

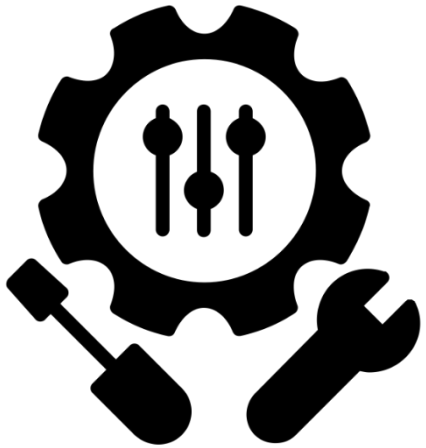
- How can we
  - explore and understand the content in a small-to-medium sized, rapidly changing collection of documents;
  - locally;
  - with RAM- and storage-constrained end-user computers;
  - with little or no coding experience needed?
- Based on frequentist text analytics and Latent Dirichlet Allocation (LDA) topic modeling conducted by STPI over the past decade



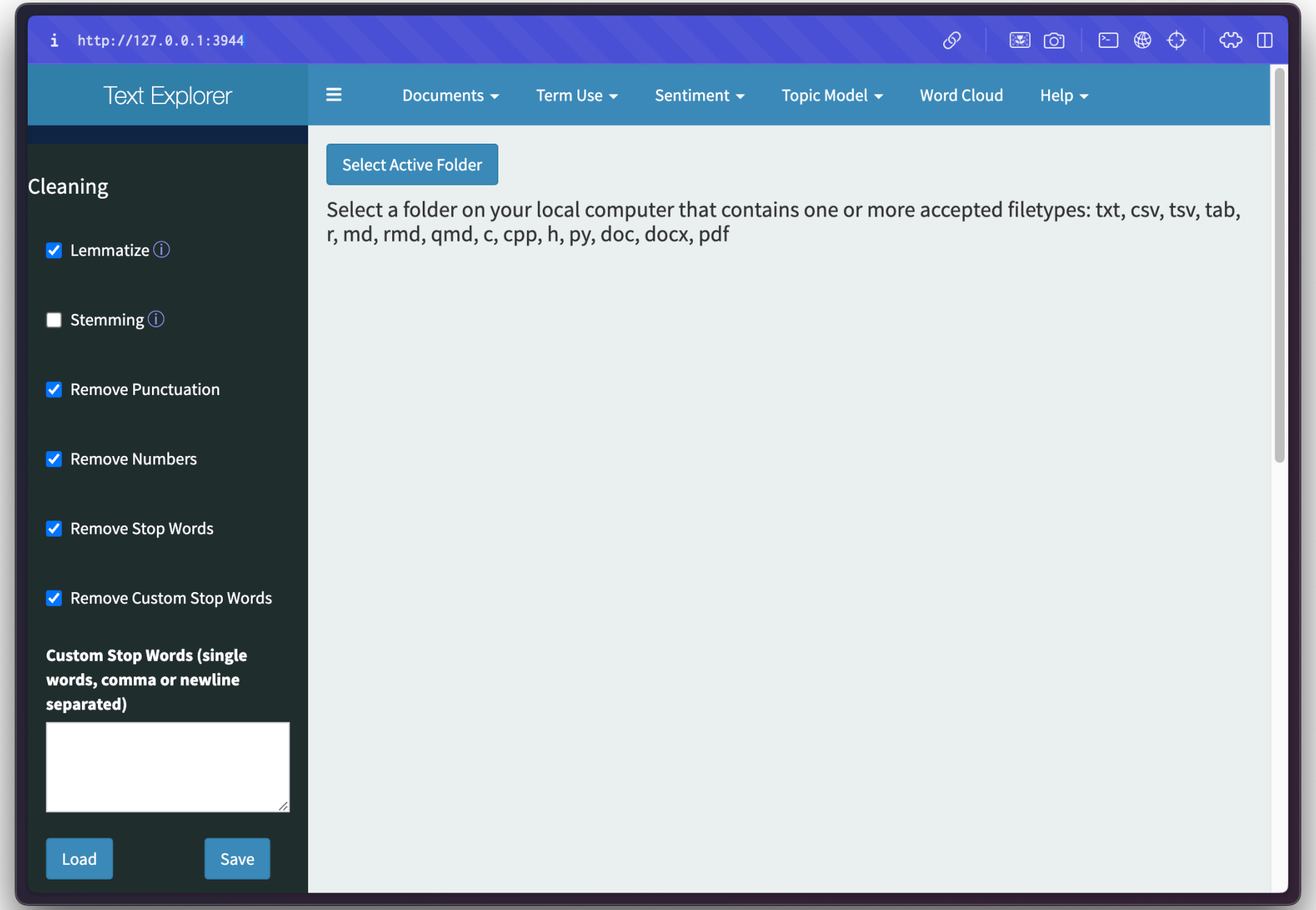
Created by UnikonSolution
from Noun Project

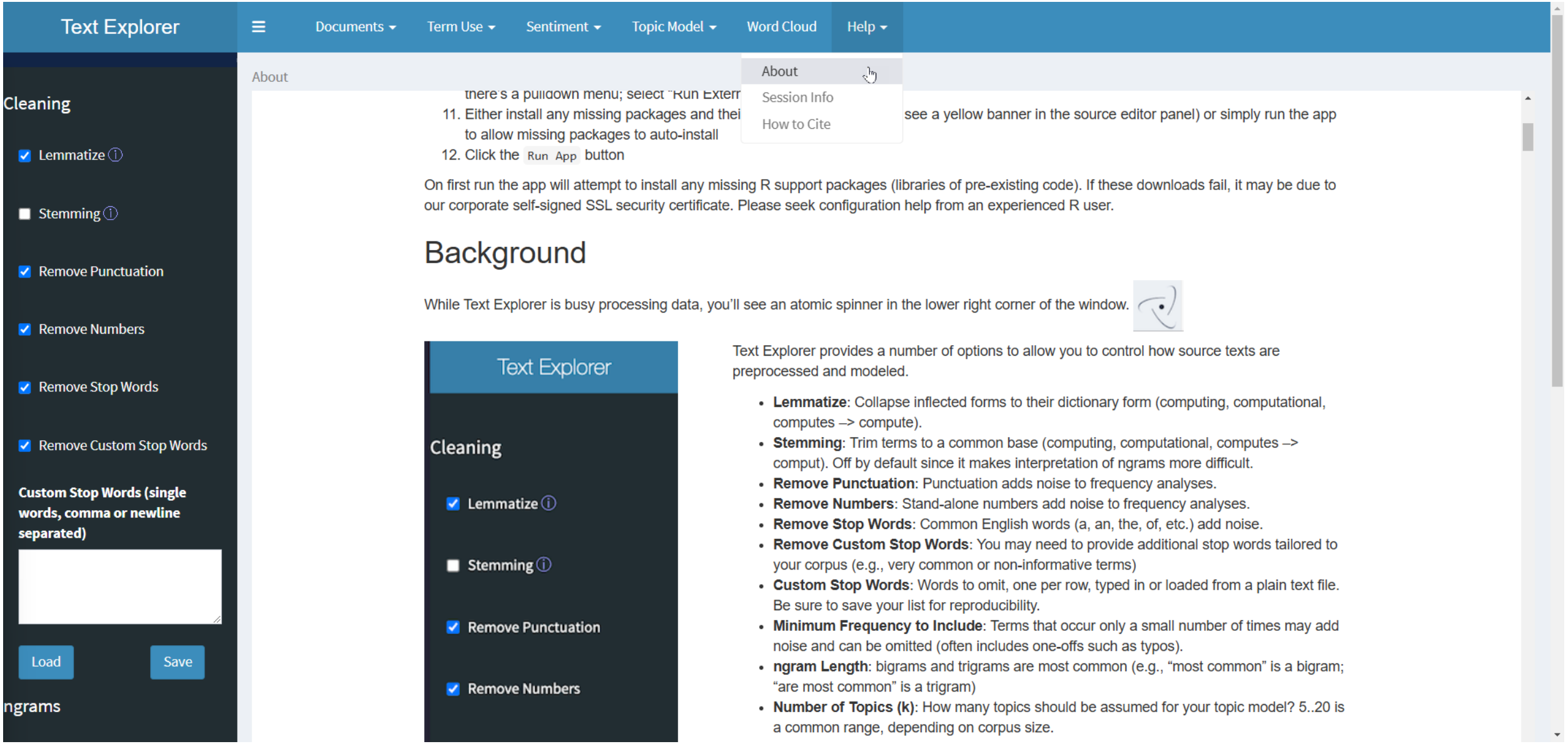
What Our Users Need to do to Get Started

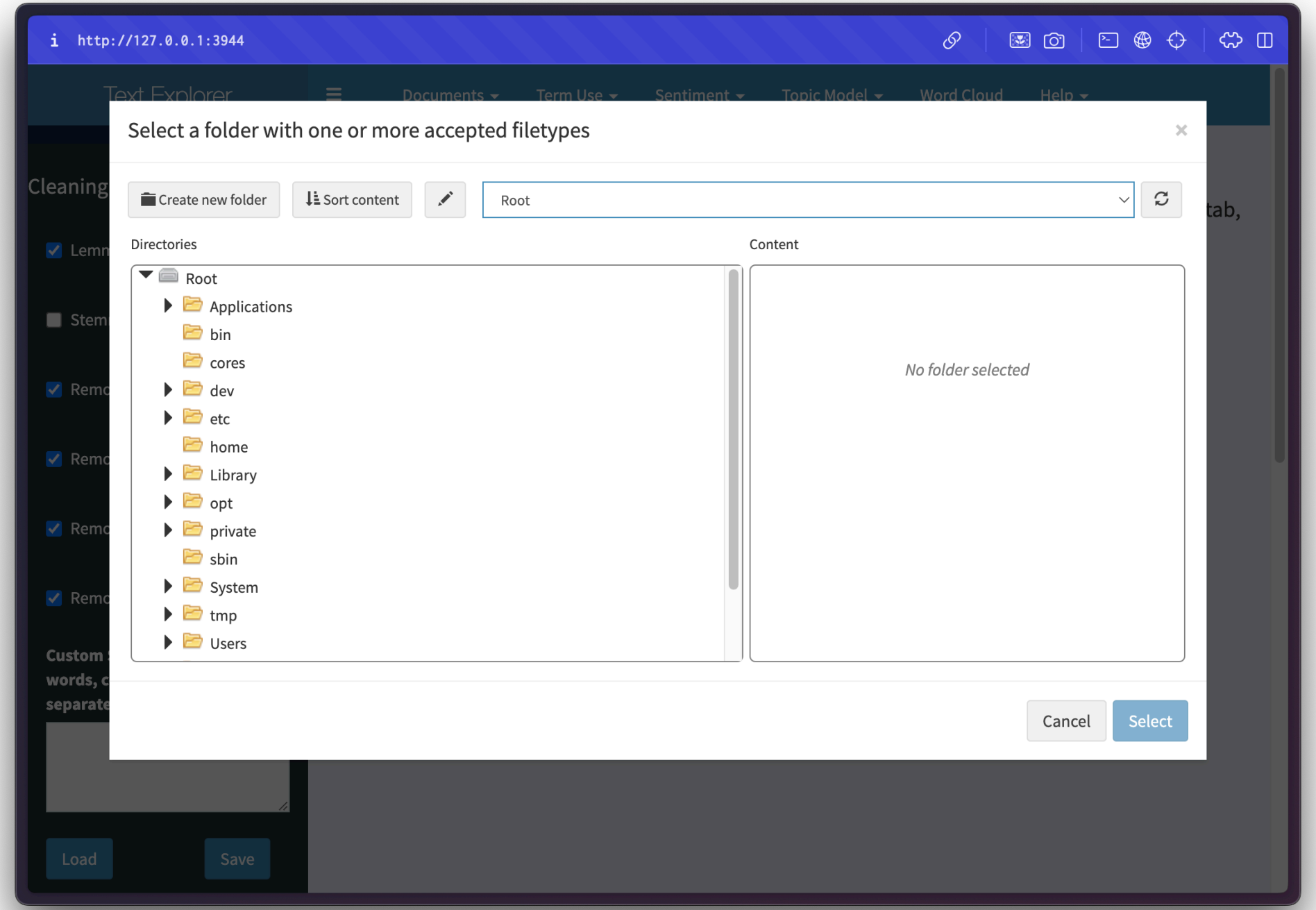
- Download and install
  - R
  - RStudio
- Download a zipped RStudio project from an enterprise code repository
- Run the Shiny app
- Select a locally-accessible document(s) folder

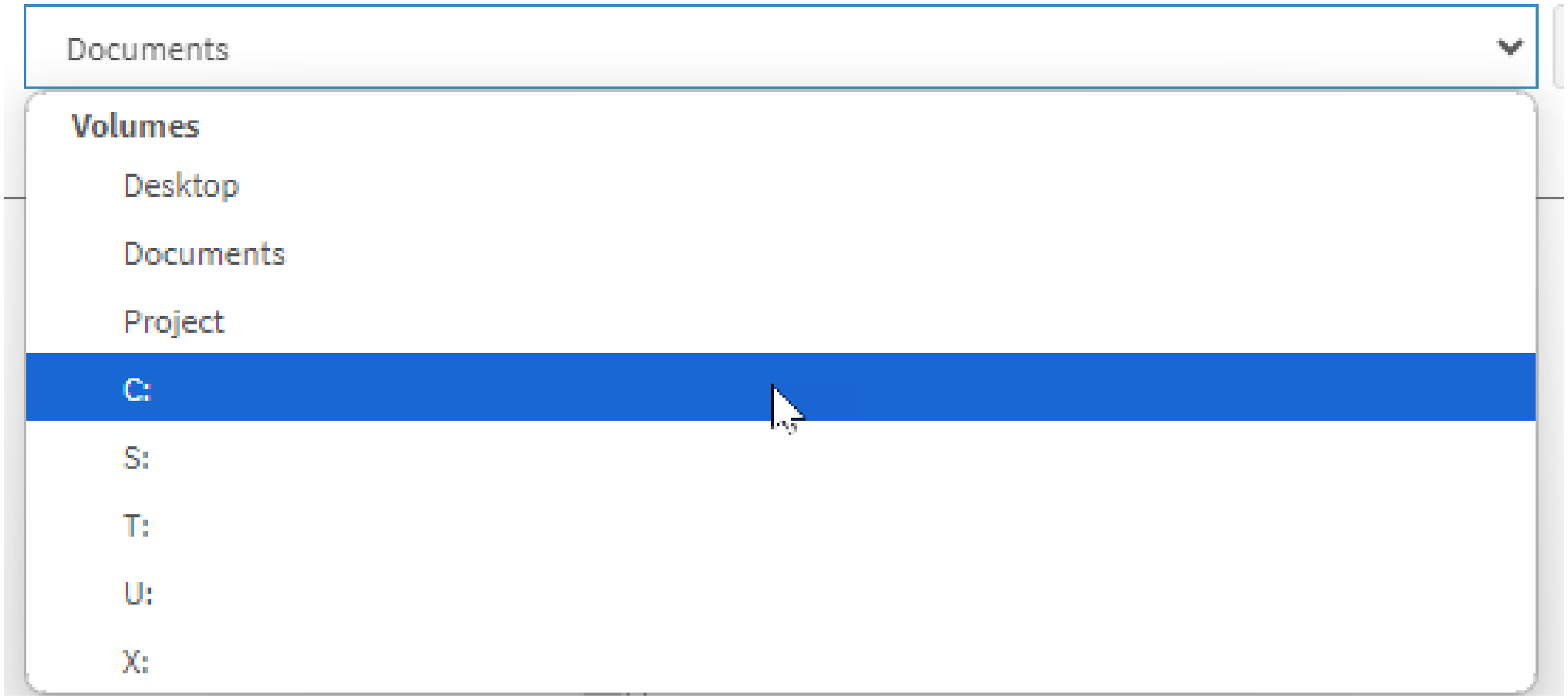


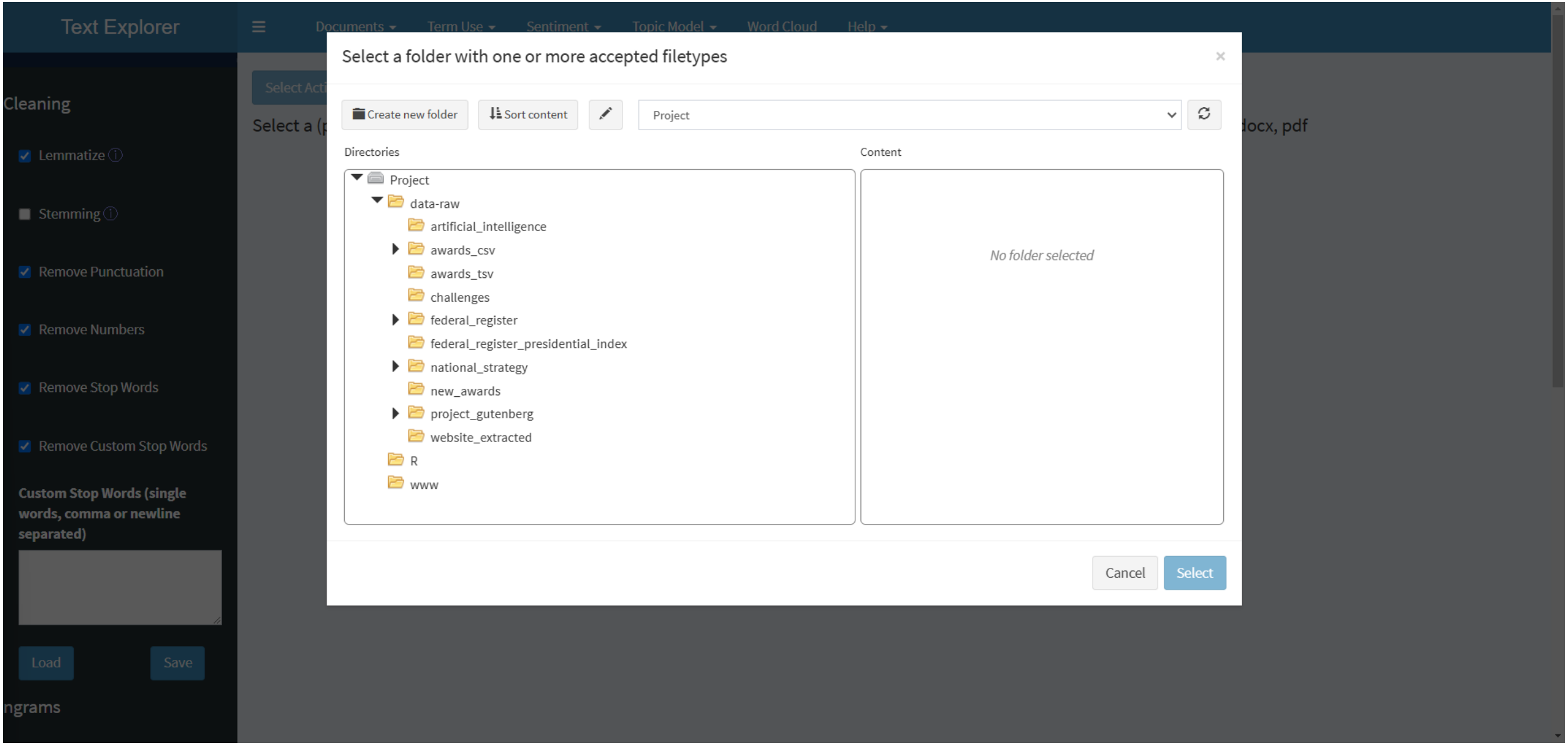
Created by fikri bayu
from Noun Project

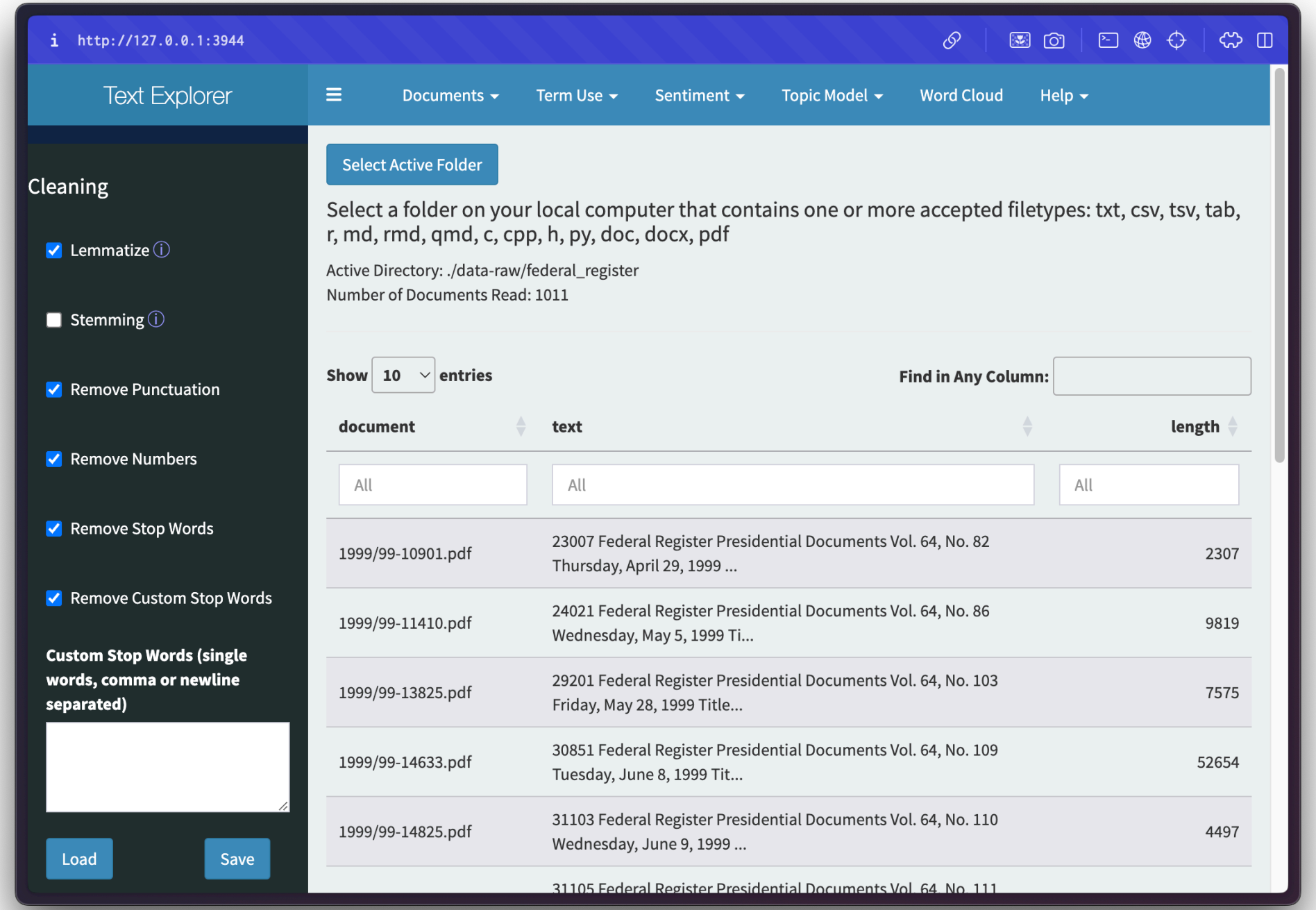

http://127.0.0.1:3944
Text Explorer
Documents
Term Use
Sentiment
Topic Model
Word Cloud
Help
Cleaning
Lemmatize
Stemming
Remove Punctuation
Remove Numbers
Remove Stop Words
Remove Custom Stop Words
Custom Stop Words (single words, comma or newline separated)
Load
Save
Select Active Folder
Select a folder on your local computer that contains one or more accepted filetypes: txt, csv, tsv, tab, r, md, rmd, qmd, c, cpp, h, py, doc, docx, pdf

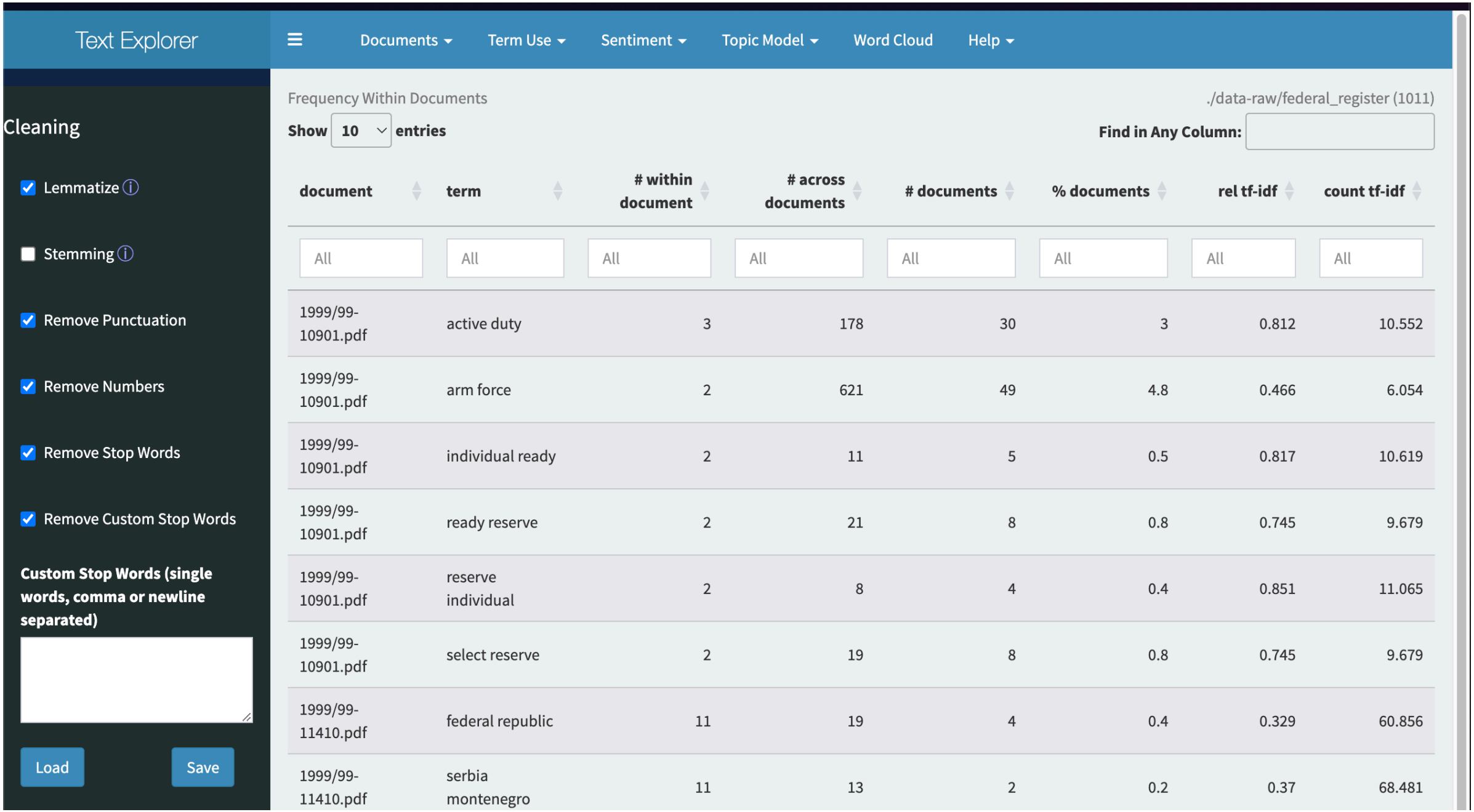

Text Explorer
Documents
Term Use
Sentiment
Topic Model
Word Cloud
Help
About
Session Info
How to Cite
About
Cleaning
Lemmatize
Stemming
Remove Punctuation
Remove Numbers
Remove Stop Words
Remove Custom Stop Words
Custom Stop Words (single words, comma or newline separated)
Load
Save
ngrams
there's a pulldown menu; select "Run Extern
11. Either install any missing packages and thei
see a yellow banner in the source editor panel) or simply run the app to allow missing packages to auto-install
12. Click the Run App button
On first run the app will attempt to install any missing R support packages (libraries of pre-existing code). If these downloads fail, it may be due to our corporate self-signed SSL security certificate. Please seek configuration help from an experienced R user.
Background
While Text Explorer is busy processing data, you'll see an atomic spinner in the lower right corner of the window.
Text Explorer
Cleaning
Lemmatize
Stemming
Remove Punctuation
Remove Numbers
Text Explorer provides a number of options to allow you to control how source texts are preprocessed and modeled.
Lemmatize: Collapse inflected forms to their dictionary form (computing, computational, computes –> compute).
Stemming: Trim terms to a common base (computing, computational, computes –> comput). Off by default since it makes interpretation of ngrams more difficult.
Remove Punctuation: Punctuation adds noise to frequency analyses.
Remove Numbers: Stand-alone numbers add noise to frequency analyses.
Remove Stop Words: Common English words (a, an, the, of, etc.) add noise.
Remove Custom Stop Words: You may need to provide additional stop words tailored to your corpus (e.g., very common or non-informative terms)
Custom Stop Words: Words to omit, one per row, typed in or loaded from a plain text file. Be sure to save your list for reproducibility.
Minimum Frequency to Include: Terms that occur only a small number of times may add noise and can be omitted (often includes one-offs such as typos).
ngram Length: bigrams and trigrams are most common (e.g., "most common" is a bigram; "are most common" is a trigram)
Number of Topics (k): How many topics should be assumed for your topic model? 5..20 is a common range, depending on corpus size.

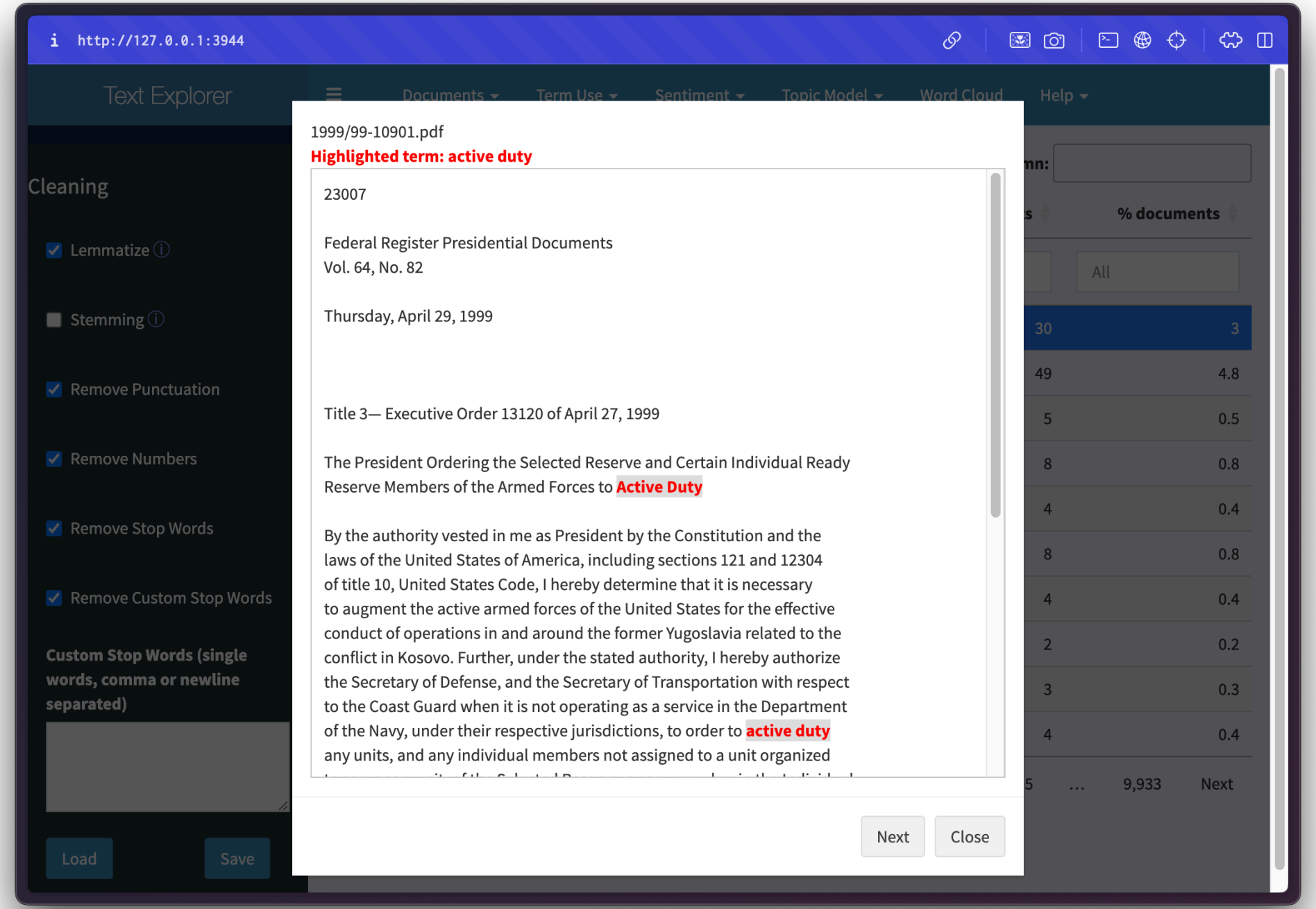

http://127.0.0.1:3944
Text Explorer
Documents
Term Use
Sentiment
Topic Model
Word Cloud
Help
Cleaning
Load
Save
Select a folder with one or more accepted filetypes
Create new folder
Sort content
Root
Directories
Content
Root
Applications
bin
cores
dev
etc
home
Library
opt
private
sbin
System
tmp
Users
No folder selected
Cancel
Select

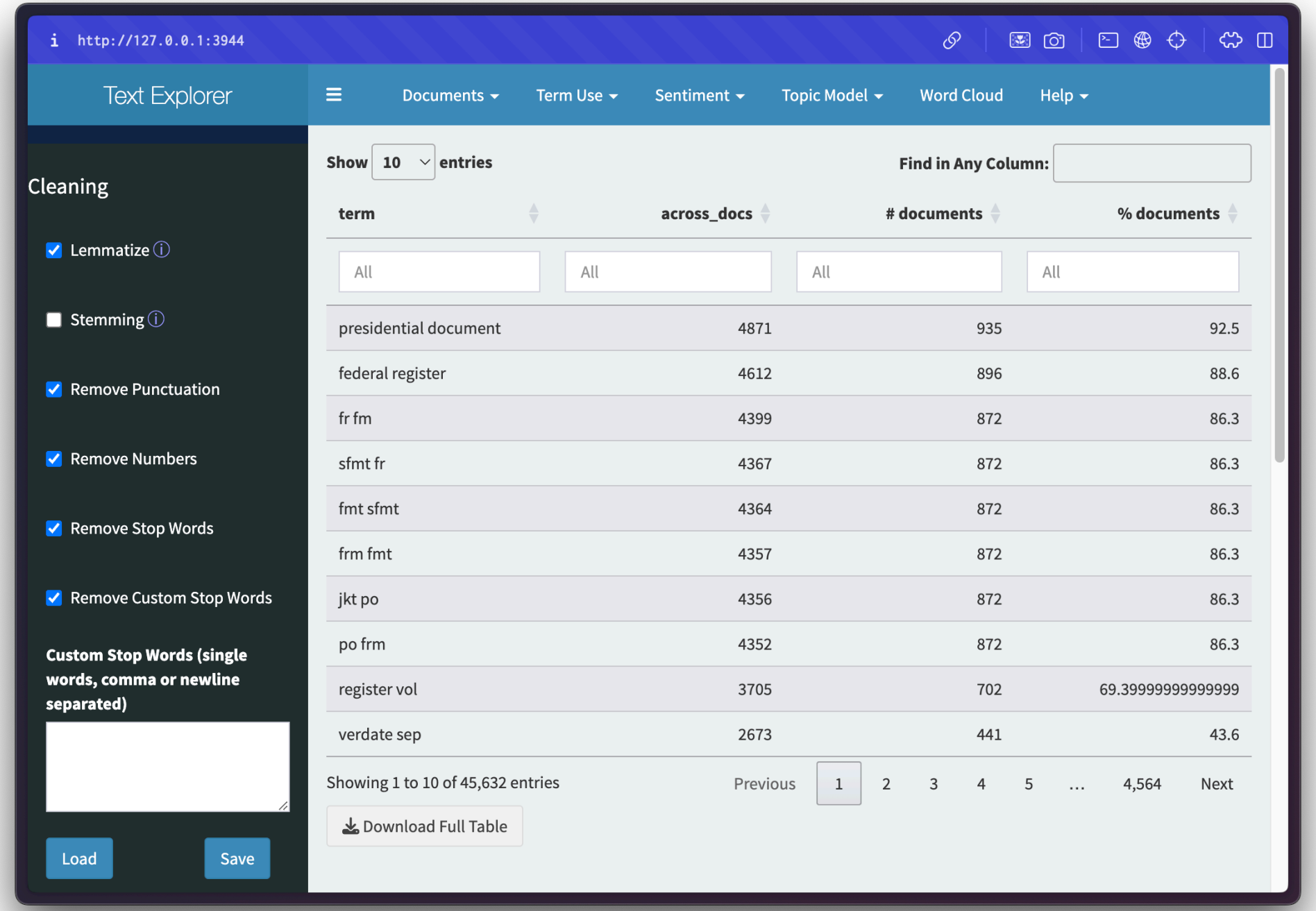

Documents
Volumes
Desktop
Documents
Project
C:
S:
T:
U:
X:

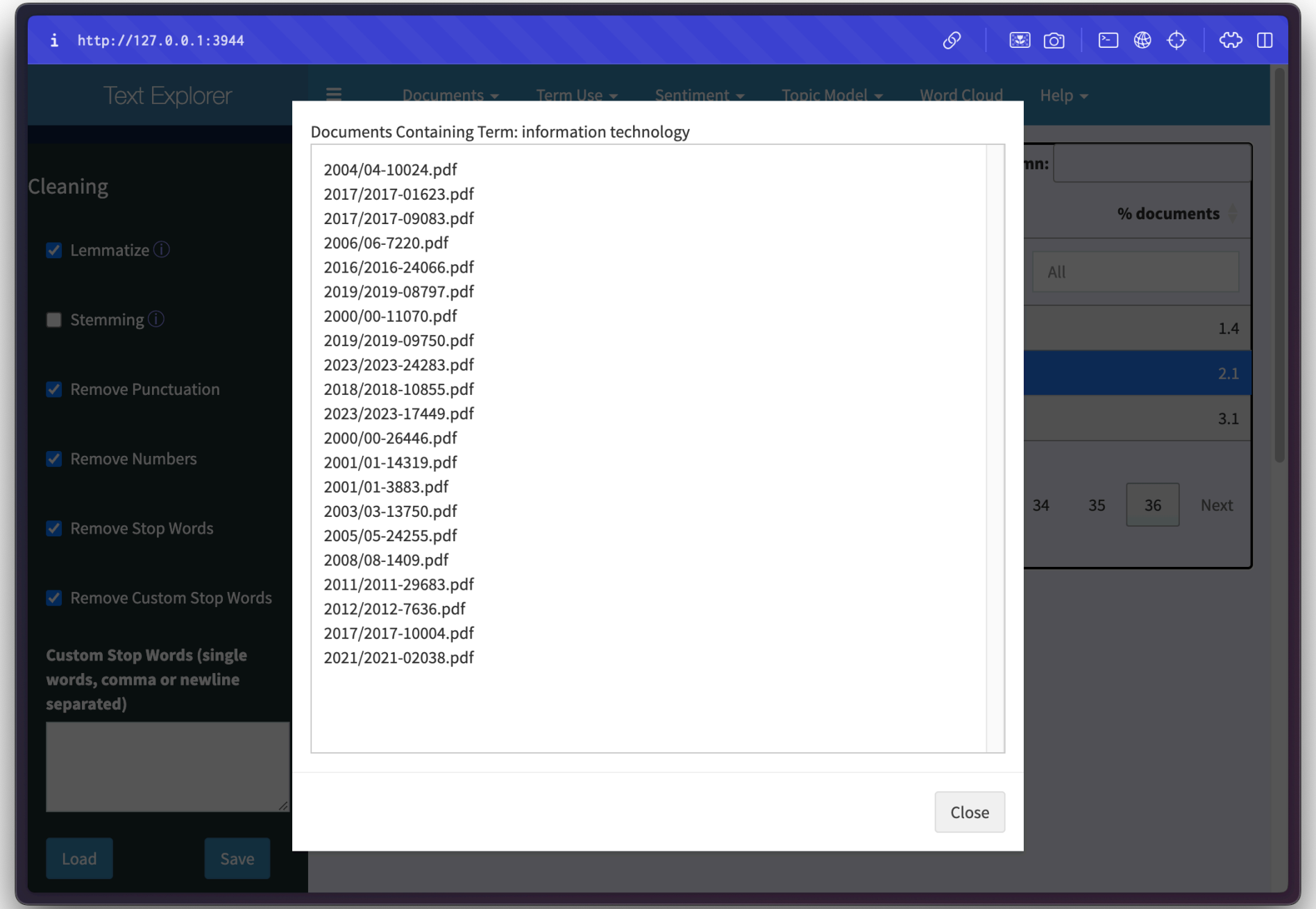

Text Explorer
Documents
Term Use
Sentiment
Topic Model
Word Cloud
Help
Cleaning
Lemmatize
Stemming
Remove Punctuation
Remove Numbers
Remove Stop Words
Remove Custom Stop Words
Custom Stop Words (single words, comma or newline separated)
Load
Save
ngrams
docx, pdf
Select a folder with one or more accepted filetypes
Create new folder
Sort content
Project
Directories
Content
Project
data-raw
artificial_intelligence
awards_csv
awards_tsv
challenges
federal_register
federal_register_presidential_index
national_strategy
new_awards
project_gutenberg
website_extracted
R
www
No folder selected
Cancel
Select

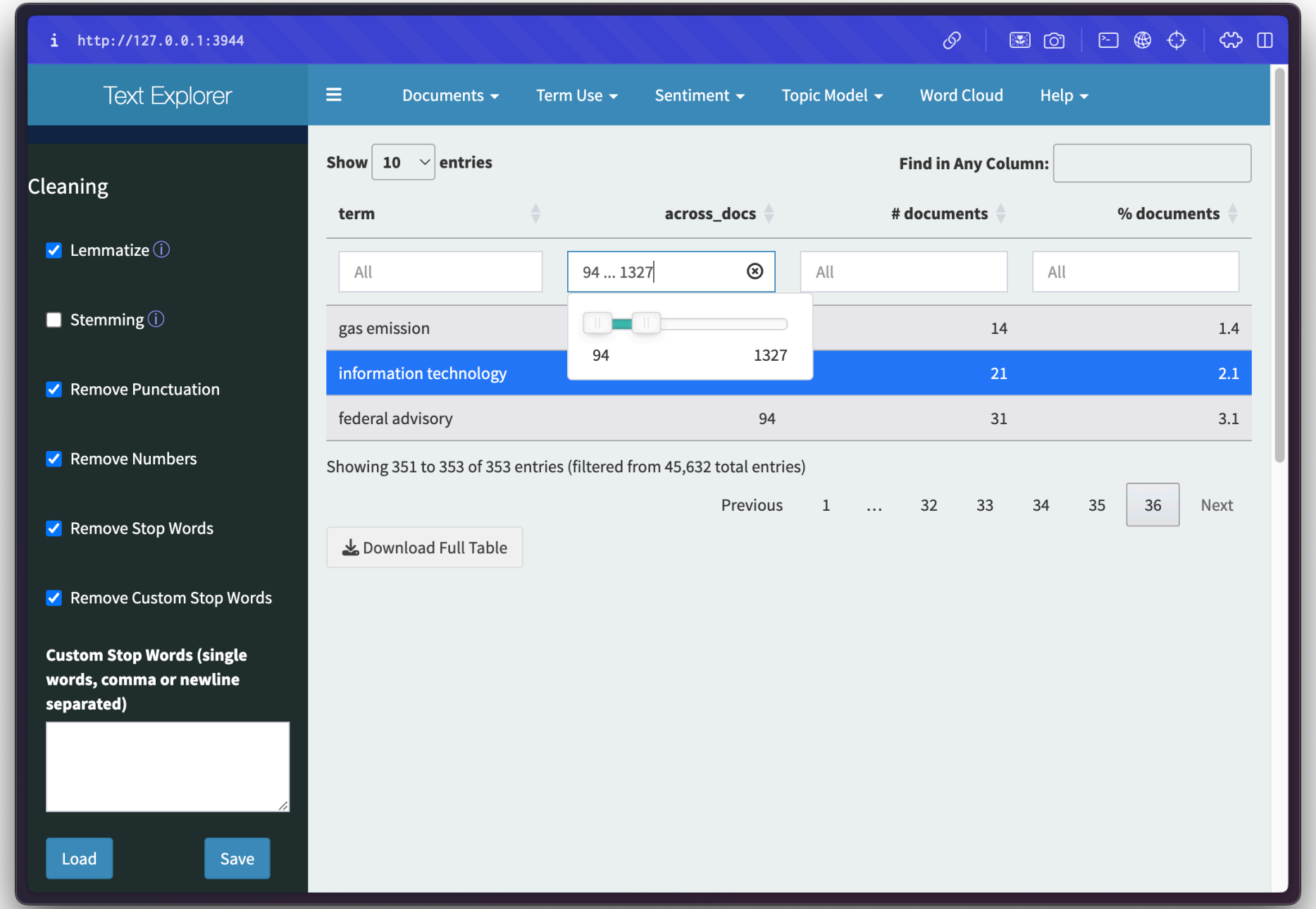

http://127.0.0.1:3944
Text Explorer
Documents
Term Use
Sentiment
Topic Model
Word Cloud
Help
Cleaning
Lemmatize
Stemming
Remove Punctuation
Remove Numbers
Remove Stop Words
Remove Custom Stop Words
Custom Stop Words (single words, comma or newline separated)
Load
Save
Select Active Folder
Select a folder on your local computer that contains one or more accepted filetypes: txt, csv, tsv, tab, r, md, rmd, qmd, c, cpp, h, py, doc, docx, pdf
Active Directory: ./data-raw/federal_register
Number of Documents Read: 1011
Show 10 entries
Find in Any Column:
document
text
length
All
All
All
1999/99-10901.pdf
23007 Federal Register Presidential Documents Vol. 64, No. 82 Thursday, April 29, 1999 ...
2307
1999/99-11410.pdf
24021 Federal Register Presidential Documents Vol. 64, No. 86 Wednesday, May 5, 1999 Ti...
9819
1999/99-13825.pdf
29201 Federal Register Presidential Documents Vol. 64, No. 103 Friday, May 28, 1999 Title...
7575
1999/99-14633.pdf
30851 Federal Register Presidential Documents Vol. 64, No. 109 Tuesday, June 8, 1999 Tit...
52654
1999/99-14825.pdf
31103 Federal Register Presidential Documents Vol. 64, No. 110 Wednesday, June 9, 1999 ...
4497

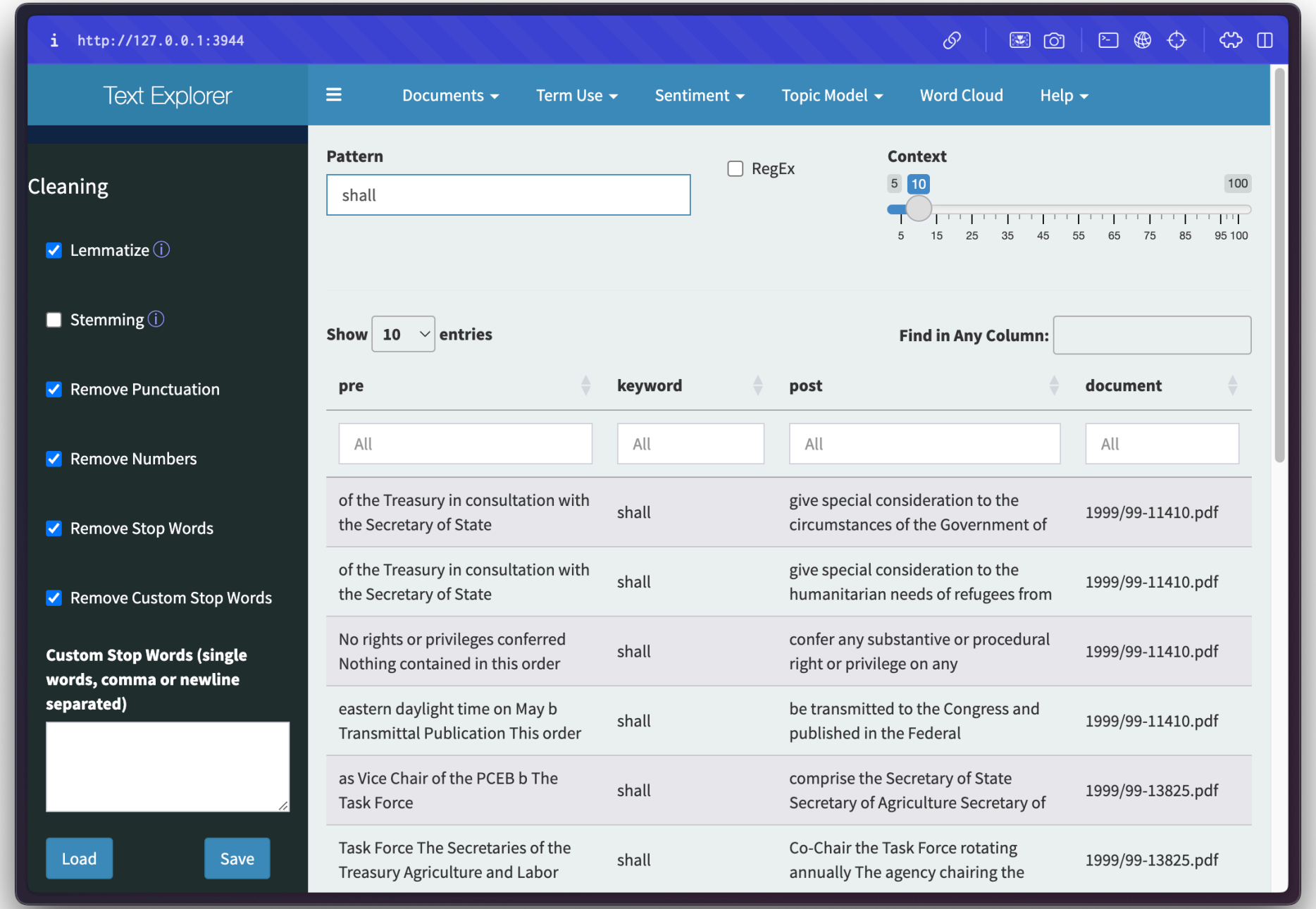

Text Explorer
Documents
Term Use
Sentiment
Topic Model
Word Cloud
Help
Cleaning
Lemmatize
Stemming
Remove Punctuation
Remove Numbers
Remove Stop Words
Remove Custom Stop Words
Custom Stop Words (single words, comma or newline separated)
Load
Save
Frequency Within Documents
./data-raw/federal_register (1011)
Show 10 entries
Find in Any Column:
document
term
within document
across documents
documents
% documents
rel tf-idf
count tf-idf
All
1999/99-10901.pdf active duty 3 178 30 3 0.812 10.552
1999/99-10901.pdf arm force 2 621 49 4.8 0.466 6.054
1999/99-10901.pdf individual ready 2 11 5 0.5 0.817 10.619
1999/99-10901.pdf ready reserve 2 21 8 0.8 0.745 9.679
1999/99-10901.pdf reserve individual 2 8 4 0.4 0.851 11.065
1999/99-10901.pdf select reserve 2 19 8 0.8 0.745 9.679
1999/99-11410.pdf federal republic 11 19 4 0.4 0.329 60.856
1999/99-11410.pdf serbia montenegro 11 13 2 0.2 0.37 68.481

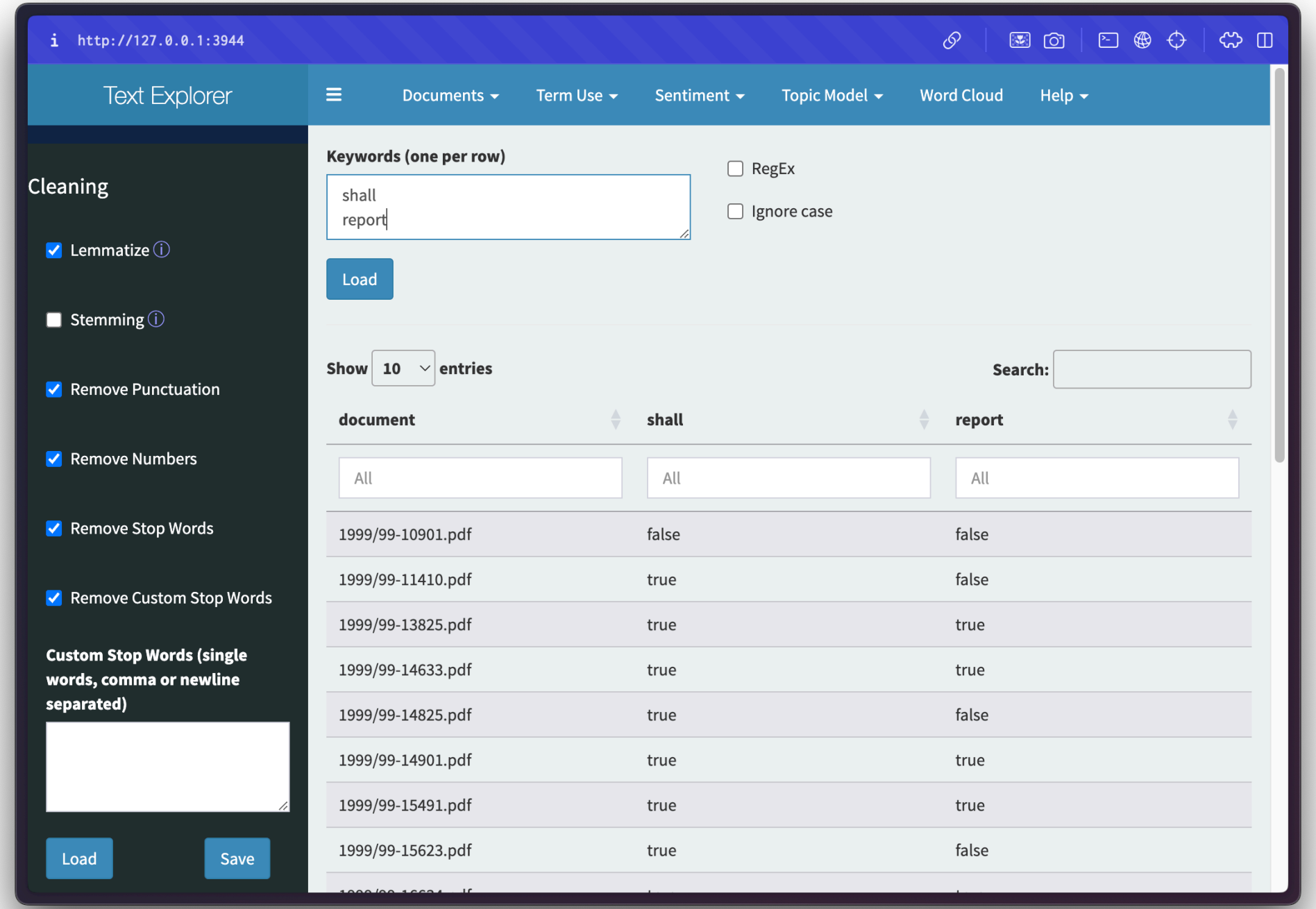

http://127.0.0.1:3944
Text Explorer
Documents
Term Use
Sentiment
Topic Model
Word Cloud
Help
Cleaning
Lemmatize
Stemming
Remove Punctuation
Remove Numbers
Remove Stop Words
Remove Custom Stop Words
Custom Stop Words (single words, comma or newline separated)
Load
Save
1999/99-10901.pdf
Highlighted term: active duty
23007
Federal Register Presidential Documents
Vol. 64, No. 82
Thursday, April 29, 1999
Title 3— Executive Order 13120 of April 27, 1999
The President Ordering the Selected Reserve and Certain Individual Ready Reserve Members of the Armed Forces to Active Duty
By the authority vested in me as President by the Constitution and the laws of the United States of America, including sections 121 and 12304 of title 10, United States Code, I hereby determine that it is necessary to augment the active armed forces of the United States for the effective conduct of operations in and around the former Yugoslavia related to the conflict in Kosovo. Further, under the stated authority, I hereby authorize the Secretary of Defense, and the Secretary of Transportation with respect to the Coast Guard when it is not operating as a service in the Department of the Navy, under their respective jurisdictions, to order to active duty any units, and any individual members not assigned to a unit organized
Next
Close
% documents
All
30
3
49
4.8
5
0.5
8
0.4
8
0.8
4
0.4
2
0.2
3
0.3
4
0.4
9,933
Next

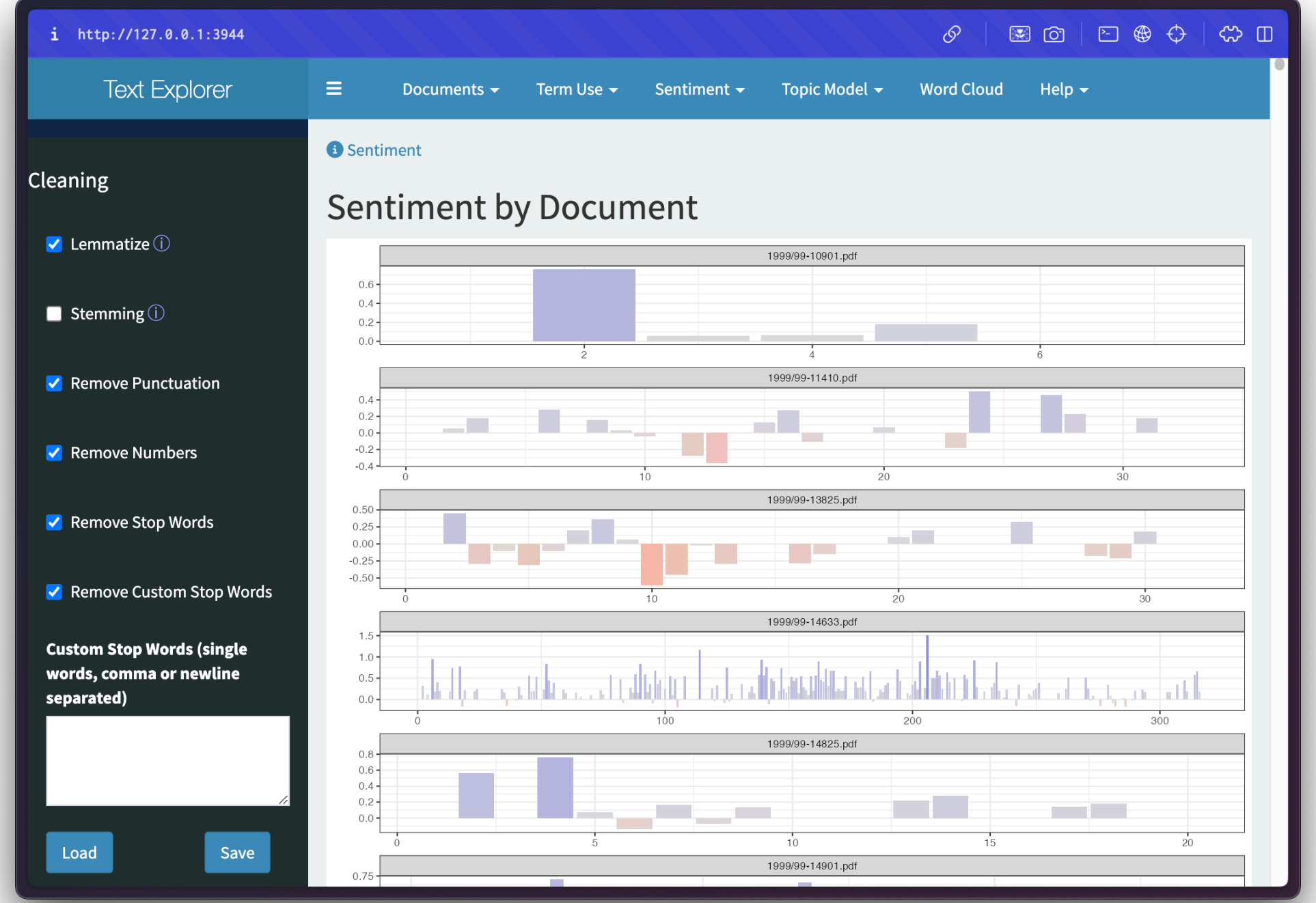

http://127.0.0.1:3944
Text Explorer
Documents
Term Use
Sentiment
Topic Model
Word Cloud
Help
Cleaning
Lemmatize
Stemming
Remove Punctuation
Remove Numbers
Remove Stop Words
Remove Custom Stop Words
Custom Stop Words (single words, comma or newline separated)
Load
Save
Show 10 entries
Find in Any Column:
term
across_docs
documents
% documents
All
presidential document 4871 935 92.5
federal register 4612 896 88.6
fr fm 4399 872 86.3
sfmt fr 4367 872 86.3
fmt sfmt 4364 872 86.3
frm fmt 4357 872 86.3
jkt po 4356 872 86.3
po frm 4352 872 86.3
register vol 3705 702 69.39999999999999
verdate sep 2673 441 43.6
Showing 1 to 10 of 45,632 entries
Previous 1 2 3 4 5 … 4,564 Next
Download Full Table

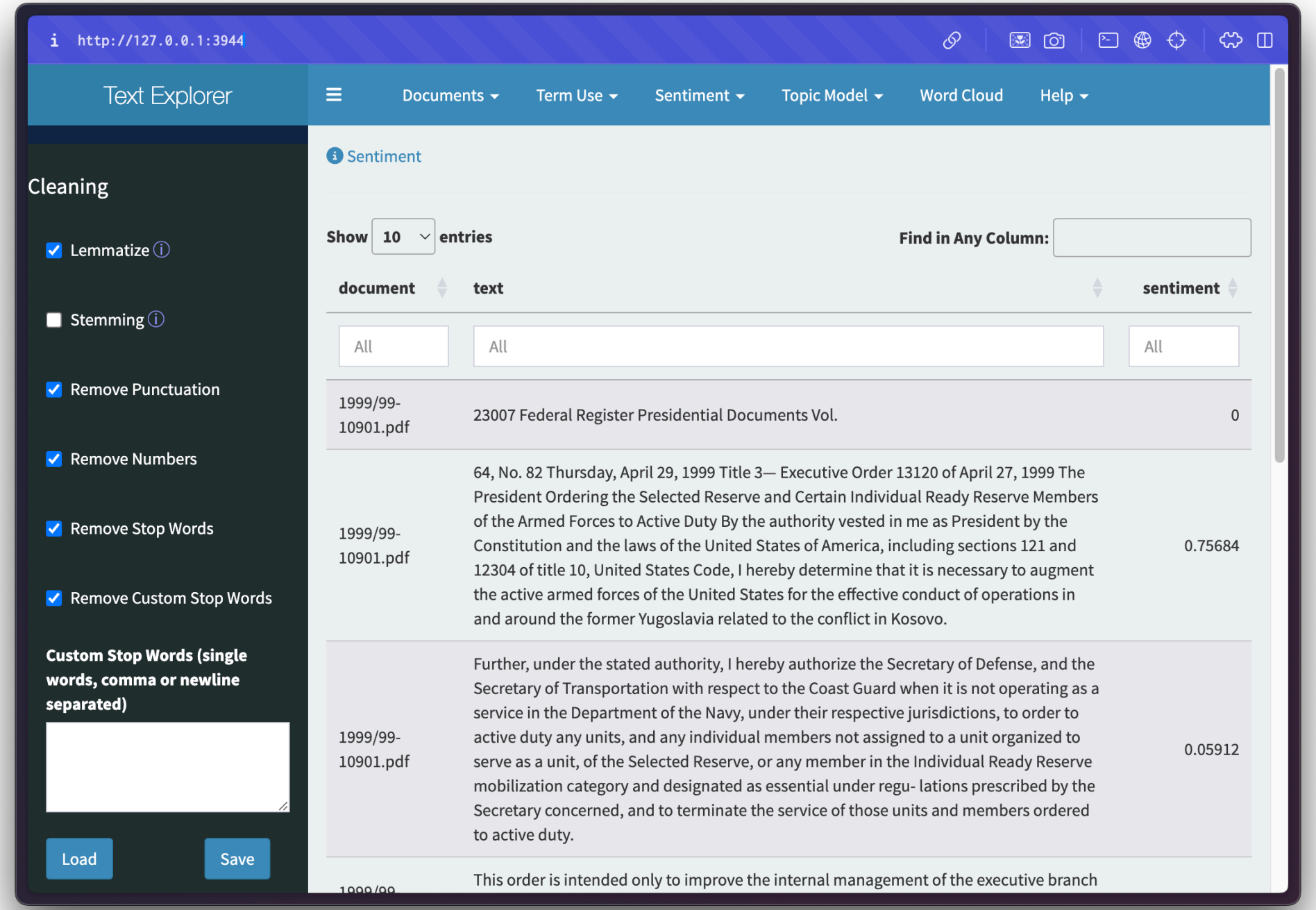

http://127.0.0.1:3944
Text Explorer
Documents
Term Use
Sentiment
Topic Model
Word Cloud
Help
Cleaning
Lemmatize
Stemming
Remove Punctuation
Remove Numbers
Remove Stop Words
Remove Custom Stop Words
Custom Stop Words (single words, comma or newline separated)
Load
Save
Documents Containing Term: information technology
2004/04-10024.pdf
2017/2017-01623.pdf
2017/2017-09083.pdf
2006/06-7220.pdf
2016/2016-24066.pdf
2019/2019-08797.pdf
2000/00-11070.pdf
2019/2019-09750.pdf
2023/2023-24283.pdf
2018/2018-10855.pdf
2023/2023-17449.pdf
2000/00-26446.pdf
2001/01-14319.pdf
2001/01-3883.pdf
2003/03-13750.pdf
2005/05-24255.pdf
2008/08-1409.pdf
2011/2011-29683.pdf
2012/2012-7636.pdf
2017/2017-10004.pdf
2021/2021-02038.pdf
Close
% documents
All
1.4
2.1
3.1
34
35
36
Next

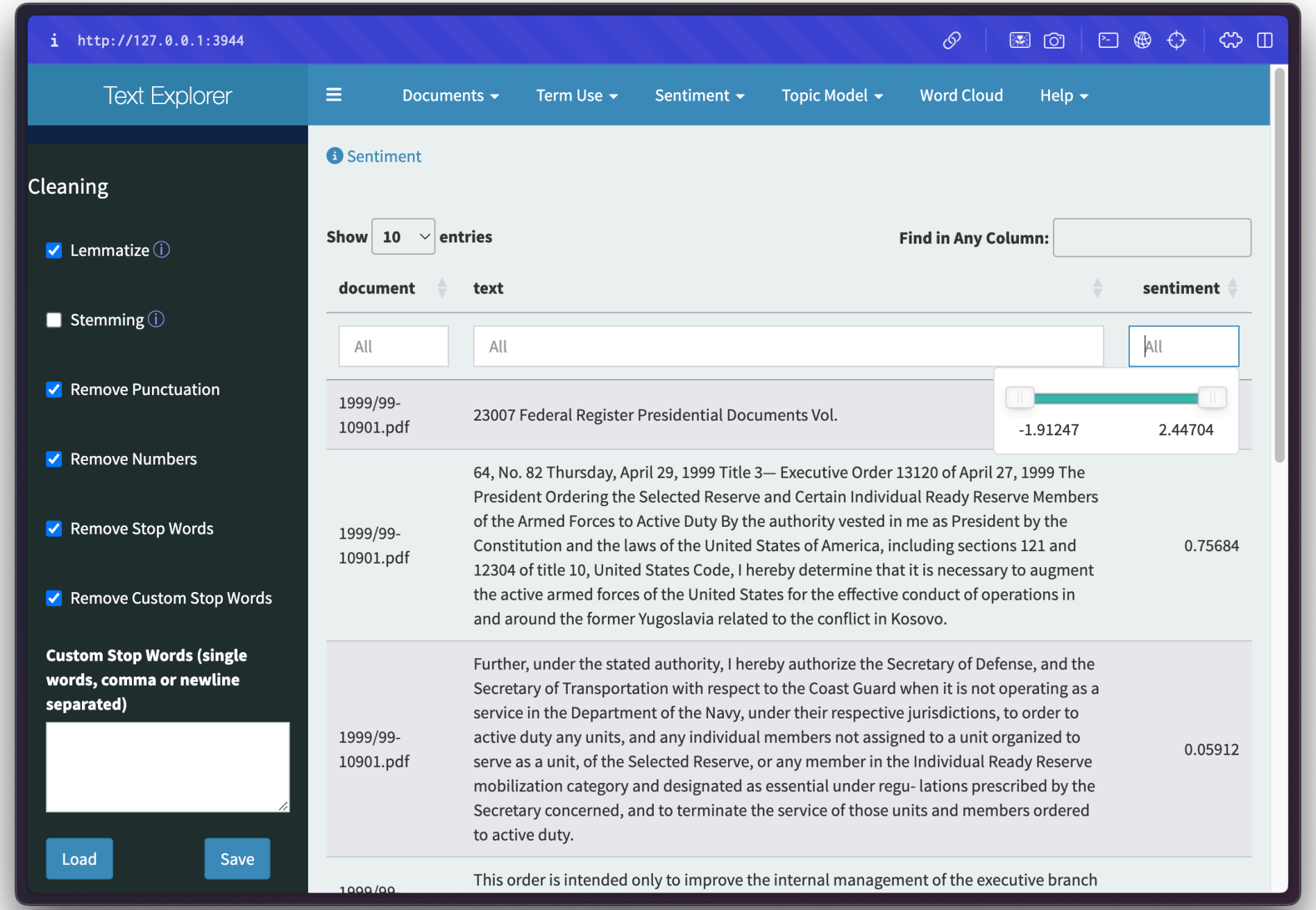

http://127.0.0.1:3944
Text Explorer
Documents
Term Use
Sentiment
Topic Model
Word Cloud
Help
Cleaning
Lemmatize
Stemming
Remove Punctuation
Remove Numbers
Remove Stop Words
Remove Custom Stop Words
Custom Stop Words (single words, comma or newline separated)
Load
Save
Show 10 entries
Find in Any Column:
term
across_docs
documents
% documents
All
94 ... 1327
All
All
94
1327
gas emission
14
1.4
information technology
21
2.1
federal advisory
94
31
3.1
Showing 351 to 353 of 353 entries (filtered from 45,632 total entries)
Previous
1
...
32
33
34
35
36
Next
Download Full Table

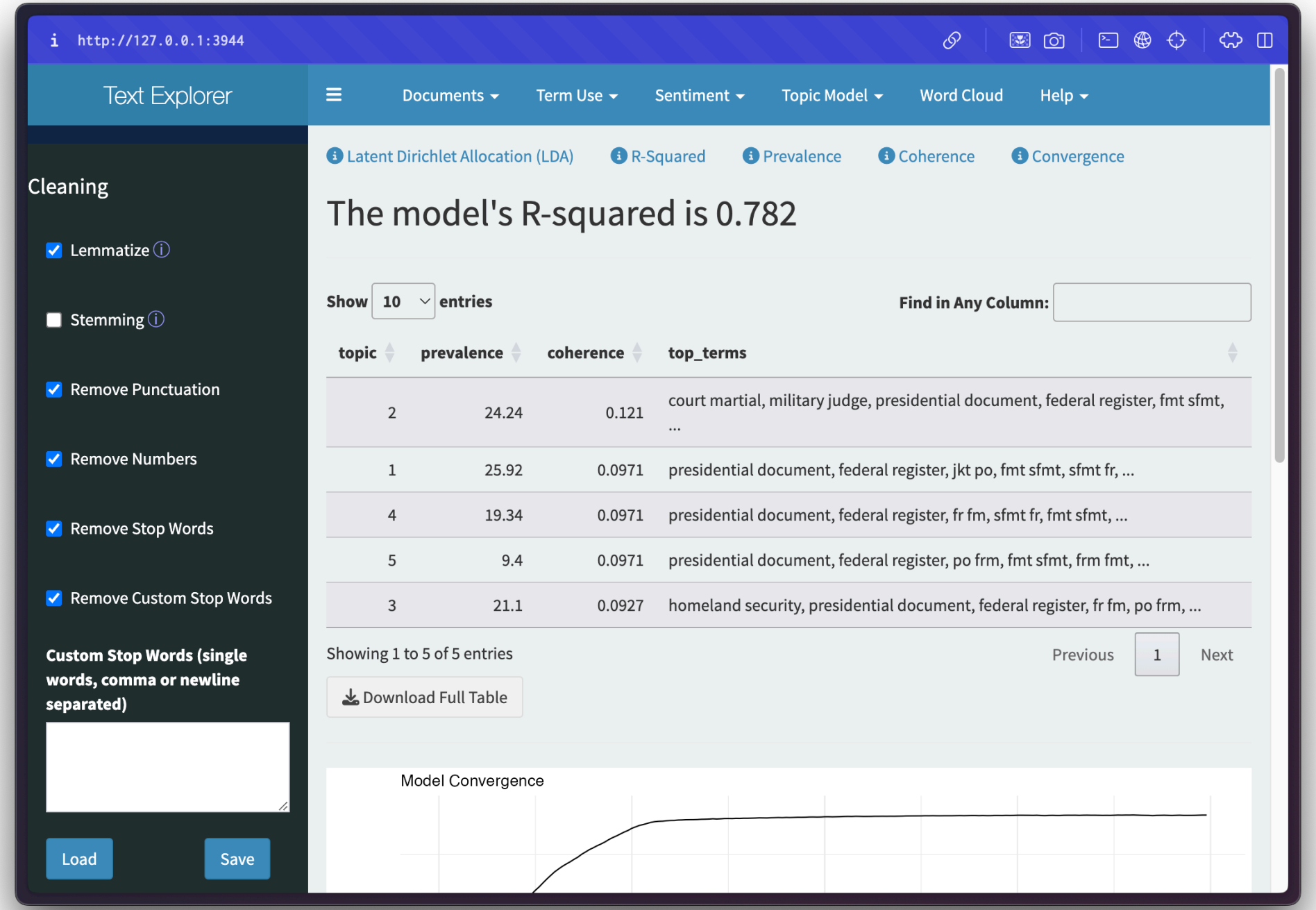

http://127.0.0.1:3944
Text Explorer
Documents
Term Use
Sentiment
Topic Model
Word Cloud
Help
Cleaning
Lemmatize
Stemming
Remove Punctuation
Remove Numbers
Remove Stop Words
Remove Custom Stop Words
Custom Stop Words (single words, comma or newline separated)
Load
Save
Pattern
shall
RegEx
Context
5
10
100
5 15 25 35 45 55 65 75 85 95 100
Show 10 entries
Find in Any Column:
pre
keyword
post
document
All
All
All
All
of the Treasury in consultation with the Secretary of State
shall
give special consideration to the circumstances of the Government of
1999/99-11410.pdf
of the Treasury in consultation with the Secretary of State
shall
give special consideration to the humanitarian needs of refugees from
1999/99-11410.pdf
No rights or privileges conferred Nothing contained in this order
shall
confer any substantive or procedural right or privilege on any
1999/99-11410.pdf
eastern daylight time on May b Transmittal Publication This order
shall
be transmitted to the Congress and published in the Federal
1999/99-11410.pdf
as Vice Chair of the PCEB b The Task Force
shall
comprise the Secretary of State Secretary of Agriculture Secretary of
1999/99-13825.pdf
Task Force The Secretaries of the Treasury Agriculture and Labor
shall
Co-Chair the Task Force rotating annually The agency chairing the
1999/99-13825.pdf


http://127.0.0.1:3944
Text Explorer
Documents
Term Use
Sentiment
Topic Model
Word Cloud
Help
Cleaning
Lemmatize
Stemming
Remove Punctuation
Remove Numbers
Remove Stop Words
Remove Custom Stop Words
Custom Stop Words (single words, comma or newline separated)
Load
Save
Keywords (one per row)
shall
report
RegEx
Ignore case
Load
Show 10 entries
Search:
document
shall
report
All
All
All
1999/99-10901.pdf false false
1999/99-11410.pdf true false
1999/99-13825.pdf true true
1999/99-14633.pdf true true
1999/99-14825.pdf true false
1999/99-14901.pdf true true
1999/99-15491.pdf true true
1999/99-15623.pdf true false


http://127.0.0.1:3944
Text Explorer
Documents
Term Use
Sentiment
Topic Model
Word Cloud
Help
Cleaning
Lemmatize
Stemming
Remove Punctuation
Remove Numbers
Remove Stop Words
Remove Custom Stop Words
Custom Stop Words (single words, comma or newline separated)
Load
Save
Sentiment
Sentiment by Document
1999/99-10901.pdf
1999/99-11410.pdf
1999/99-13825.pdf
1999/99-14633.pdf
1999/99-14825.pdf
1999/99-14901.pdf


http://127.0.0.1:3944
Text Explorer
Documents
Term Use
Sentiment
Topic Model
Word Cloud
Help
Cleaning
Lemmatize
Stemming
Remove Punctuation
Remove Numbers
Remove Stop Words
Remove Custom Stop Words
Custom Stop Words (single words, comma or newline separated)
Load
Save
Sentiment
Show 10 entries
Find in Any Column:
document
text
sentiment
All
All
All
1999/99-10901.pdf
23007 Federal Register Presidential Documents Vol.
0
1999/99-10901.pdf
64, No. 82 Thursday, April 29, 1999 Title 3— Executive Order 13120 of April 27, 1999 The President Ordering the Selected Reserve and Certain Individual Ready Reserve Members of the Armed Forces to Active Duty By the authority vested in me as President by the Constitution and the laws of the United States of America, including sections 121 and 12304 of title 10, United States Code, I hereby determine that it is necessary to augment the active armed forces of the United States for the effective conduct of operations in and around the former Yugoslavia related to the conflict in Kosovo.
0.75684
1999/99-10901.pdf
Further, under the stated authority, I hereby authorize the Secretary of Defense, and the Secretary of Transportation with respect to the Coast Guard when it is not operating as a service in the Department of the Navy, under their respective jurisdictions, to order to active duty any units, and any individual members not assigned to a unit organized to serve as a unit, of the Selected Reserve, or any member in the Individual Ready Reserve mobilization category and designated as essential under regu- lations prescribed by the Secretary concerned, and to terminate the service of those units and members ordered to active duty.
0.05912
This order is intended only to improve the internal management of the executive branch


http://127.0.0.1:3944
Text Explorer
Documents
Term Use
Sentiment
Topic Model
Word Cloud
Help
Sentiment
Cleaning
Lemmatize
Stemming
Remove Punctuation
Remove Numbers
Remove Stop Words
Remove Custom Stop Words
Custom Stop Words (single words, comma or newline separated)
Load
Save
Show 10 entries
Find in Any Column:
document
text
sentiment
All
All
All
-1.91247
2.44704
1999/99-10901.pdf
23007 Federal Register Presidential Documents Vol.
1999/99-10901.pdf
64, No. 82 Thursday, April 29, 1999 Title 3— Executive Order 13120 of April 27, 1999 The President Ordering the Selected Reserve and Certain Individual Ready Reserve Members of the Armed Forces to Active Duty By the authority vested in me as President by the Constitution and the laws of the United States of America, including sections 121 and 12304 of title 10, United States Code, I hereby determine that it is necessary to augment the active armed forces of the United States for the effective conduct of operations in and around the former Yugoslavia related to the conflict in Kosovo.
0.75684
1999/99-10901.pdf
Further, under the stated authority, I hereby authorize the Secretary of Defense, and the Secretary of Transportation with respect to the Coast Guard when it is not operating as a service in the Department of the Navy, under their respective jurisdictions, to order to active duty any units, and any individual members not assigned to a unit organized to serve as a unit, of the Selected Reserve, or any member in the Individual Ready Reserve mobilization category and designated as essential under regu- lations prescribed by the Secretary concerned, and to terminate the service of those units and members ordered to active duty.
0.05912
This order is intended only to improve the internal management of the executive branch


http://127.0.0.1:3944
Text Explorer
Documents
Term Use
Sentiment
Topic Model
Word Cloud
Help
Cleaning
Lemmatize
Stemming
Remove Punctuation
Remove Numbers
Remove Stop Words
Remove Custom Stop Words
Custom Stop Words (single words, comma or newline separated)
Load
Save
Latent Dirichlet Allocation (LDA)
R-Squared
Prevalence
Coherence
Convergence
The model's R-squared is 0.782
Show 10 entries
Find in Any Column:
topic
prevalence
coherence
top_terms
2 24.24 0.121 court martial, military judge, presidential document, federal register, fmt sfmt, ...
1 25.92 0.0971 presidential document, federal register, jkt po, fmt sfmt, sfmt fr, ...
4 19.34 0.0971 presidential document, federal register, fr fm, sfmt fr, fmt sfmt, ...
5 9.4 0.0971 presidential document, federal register, po frm, fmt sfmt, frm fmt, ...
3 21.1 0.0927 homeland security, presidential document, federal register, fr fm, po frm, ...
Showing 1 to 5 of 5 entries
Previous
1
Next
Download Full Table
Model Convergence

http://127.0.0.1:3944

Remove Custom Stop Words

Custom Stop Words (single words, comma or newline separated)

Load Save

ngrams

Minimum Frequency to Include

2 100

2 12 22 32 42 52 62 72 82 92 100

ngram Length

1 2 10

1 2 3 4 5 6 7 8 9 10

Topic Modeling

Number of Topics (k)

2 5 30

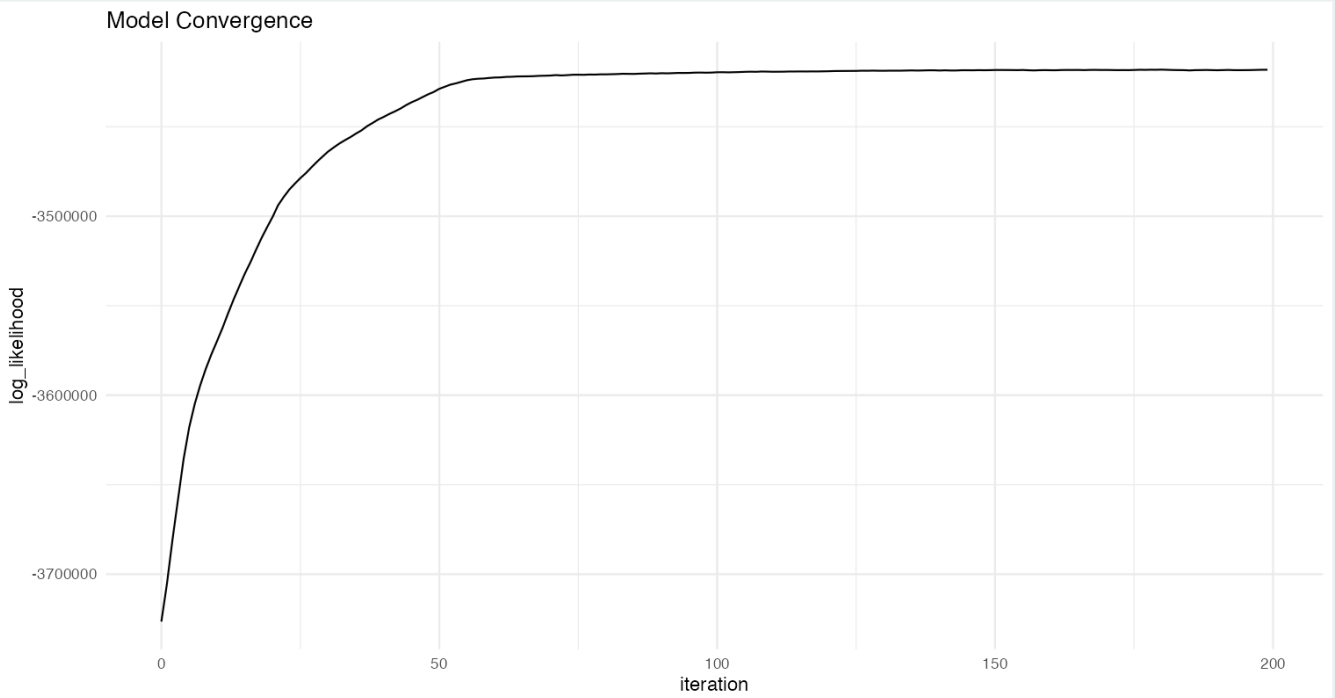
2 5 8 11 14 17 20 23 26 29 30

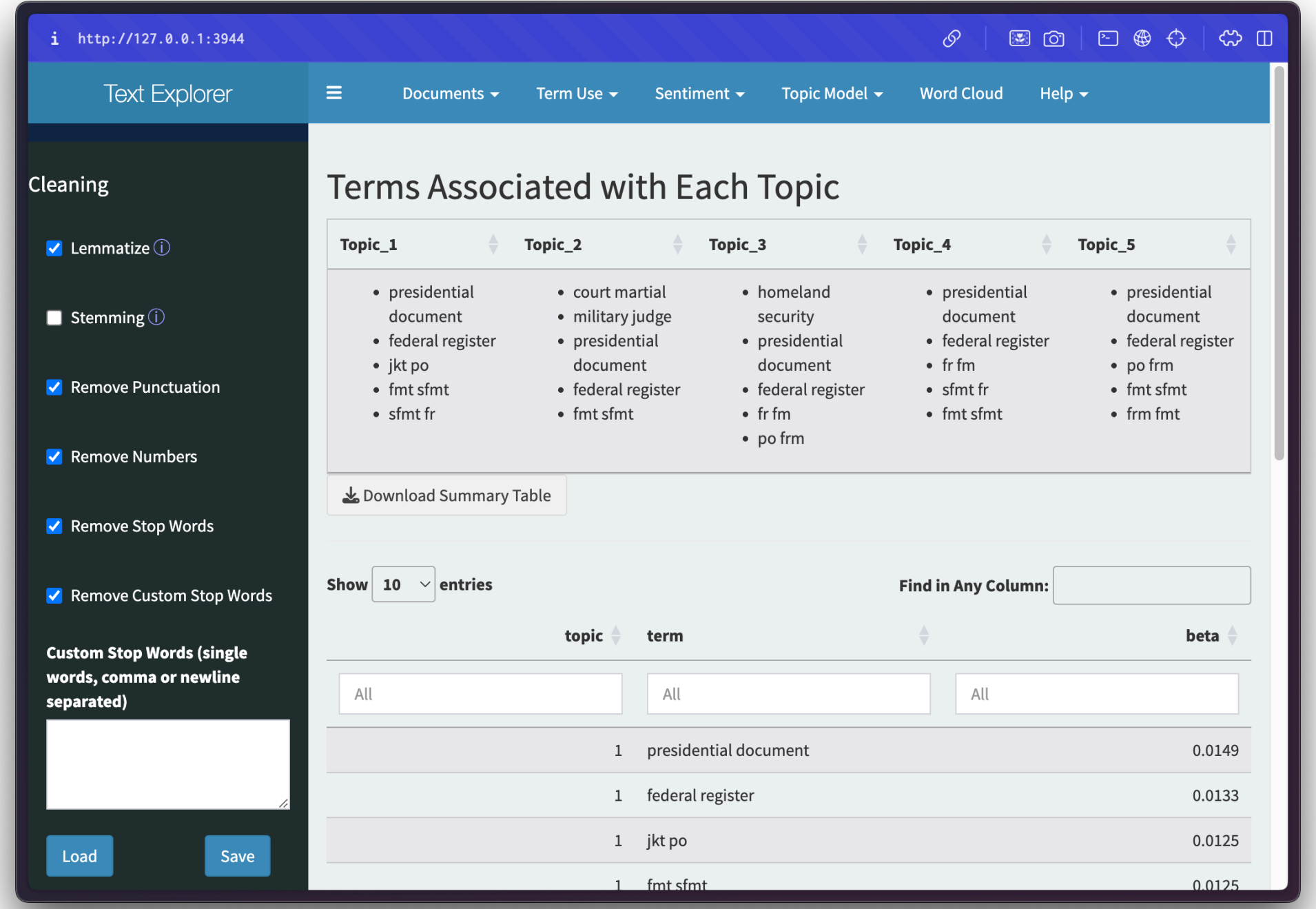
Number of Iterations

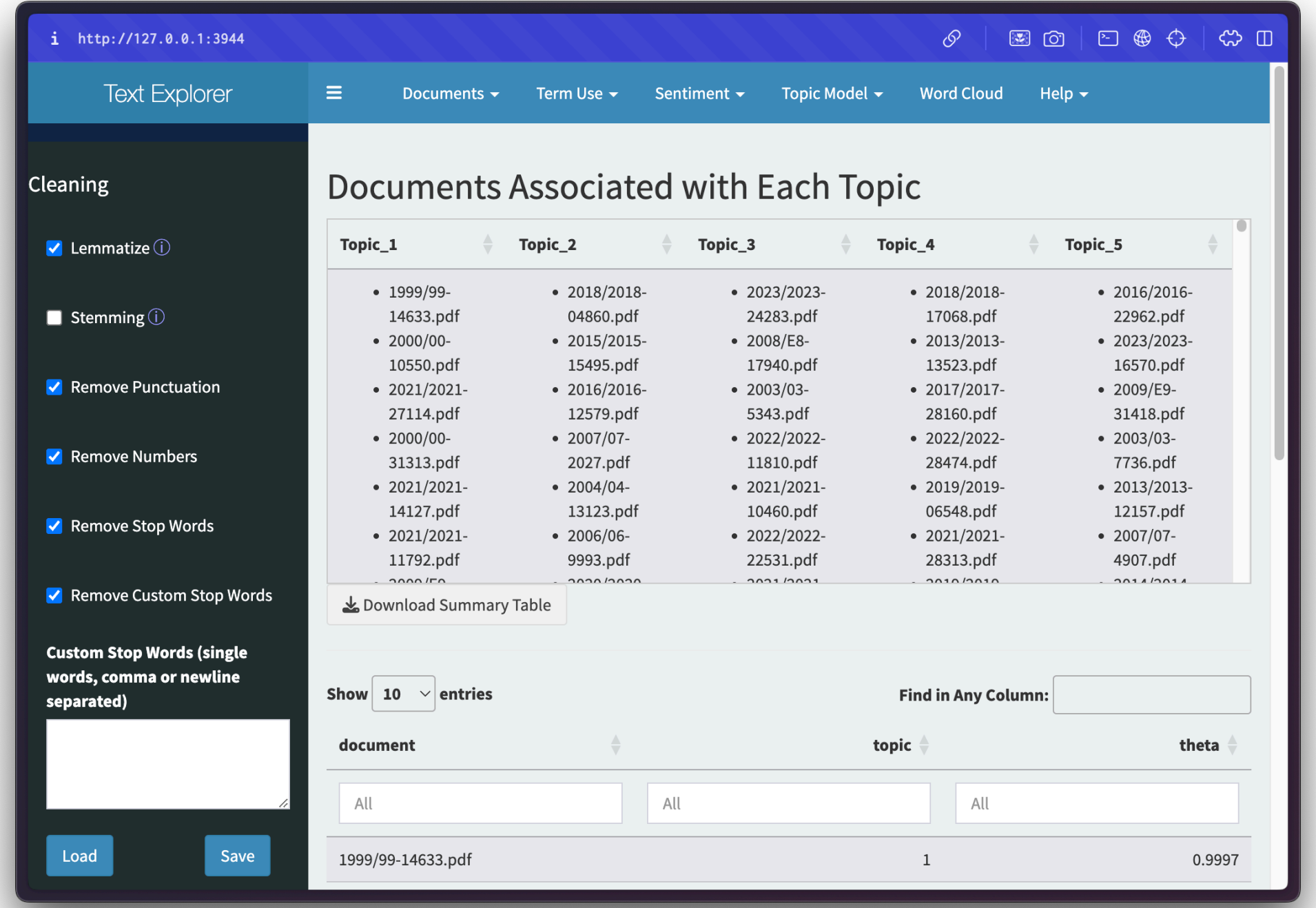
3	21.1	0.0927	homeland security, presidential document, federal register, fr fm, po frm, ...

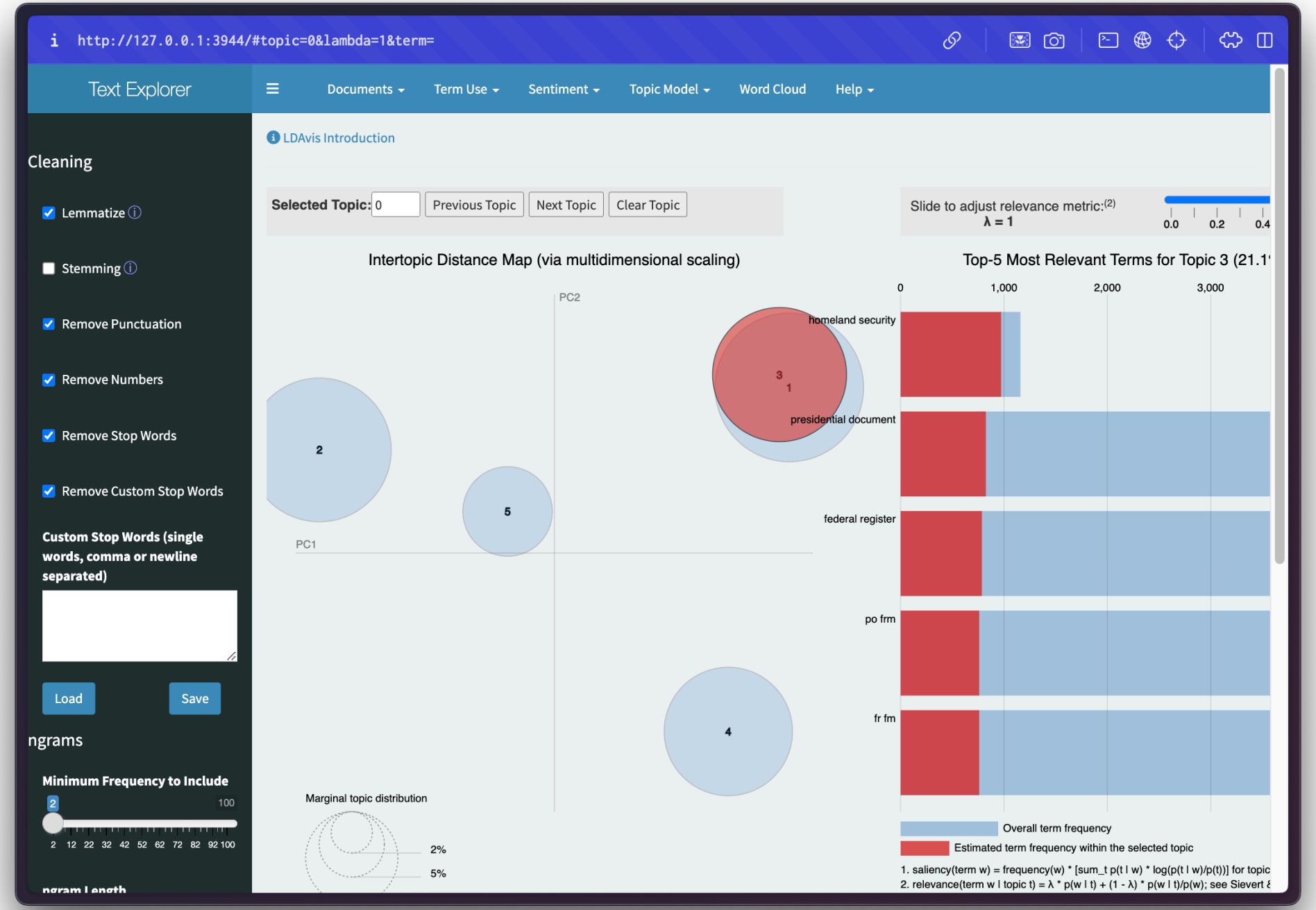
Showing 1 to 5 of 5 entries Previous 1 Next

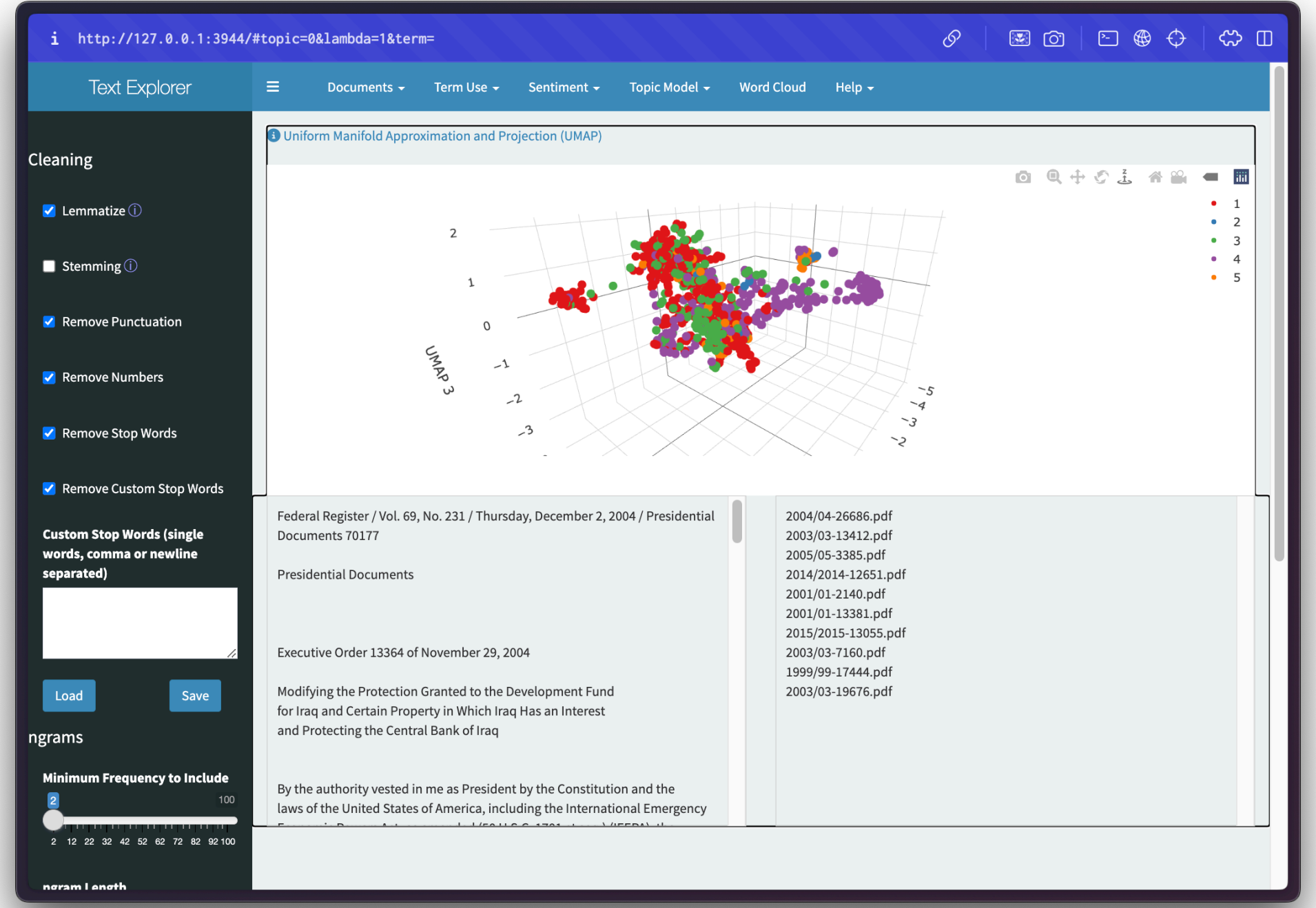
Download Full Table

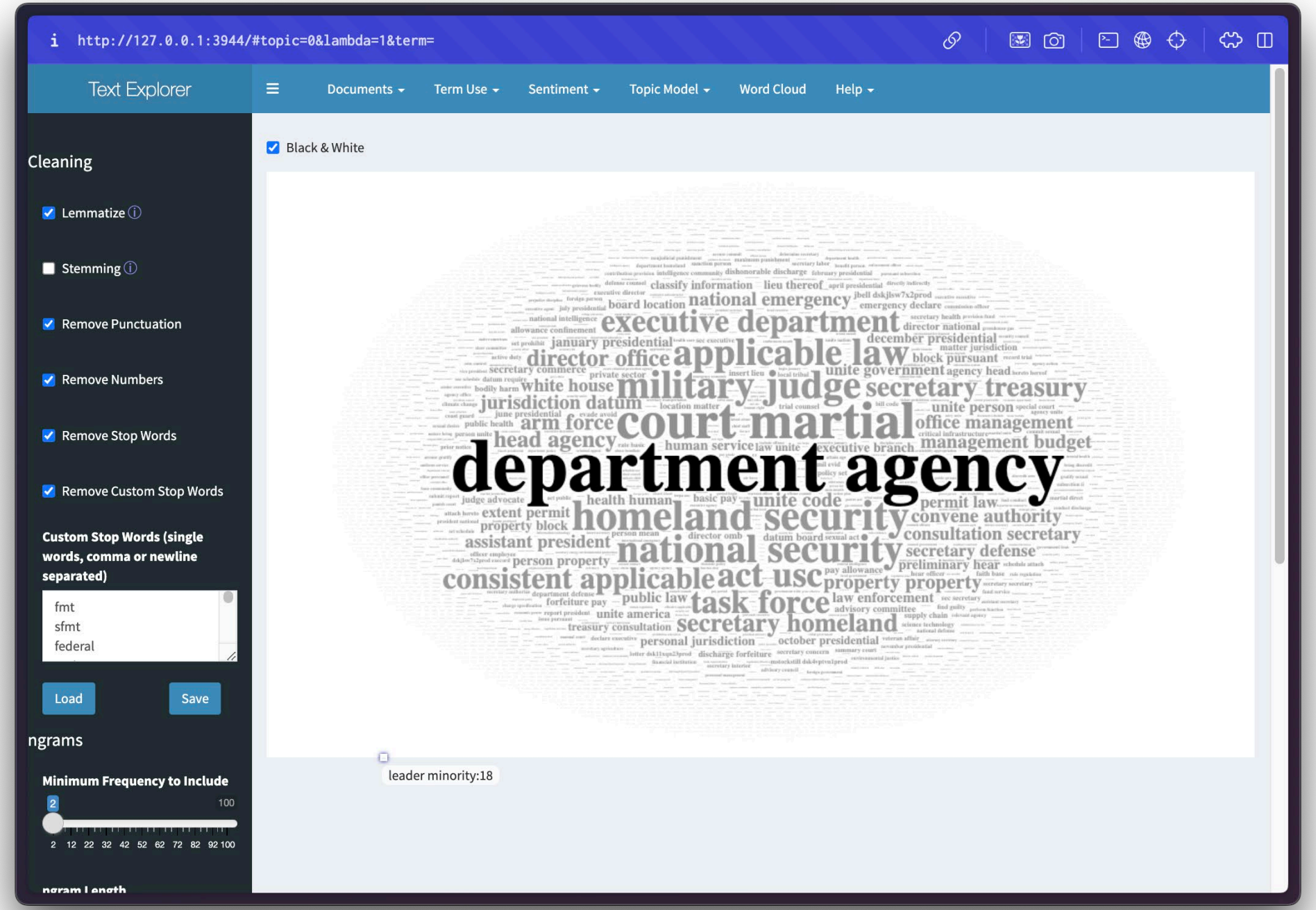



http://127.0.0.1:3944
Text Explorer
Documents
Term Use
Sentiment
Topic Model
Word Cloud
Help
Cleaning
Lemmatize
Stemming
Remove Punctuation
Remove Numbers
Remove Stop Words
Remove Custom Stop Words
Custom Stop Words (single words, comma or newline separated)
Load
Save
Terms Associated with Each Topic
Topic_1
Topic_2
Topic_3
Topic_4
Topic_5
presidential document
federal register
jkt po
fmt sfmt
sfmt fr
court martial
military judge
presidential document
federal register
fmt sfmt
homeland security
presidential document
federal register
fr fm
po frm
presidential document
federal register
fr fm
sfmt fr
fmt sfmt
presidential document
federal register
po frm
fmt sfmt
frm fmt
Download Summary Table
Show 10 entries
Find in Any Column:
topic
term
beta
All
All
All
1 presidential document 0.0149
1 federal register 0.0133
1 jkt po 0.0125
1 fmt sfmt 0.0125


http://127.0.0.1:3944
Text Explorer
Documents
Term Use
Sentiment
Topic Model
Word Cloud
Help
Cleaning
Lemmatize
Stemming
Remove Punctuation
Remove Numbers
Remove Stop Words
Remove Custom Stop Words
Custom Stop Words (single words, comma or newline separated)
Load
Save
Documents Associated with Each Topic
Topic_1
Topic_2
Topic_3
Topic_4
Topic_5
1999/99-14633.pdf
2000/00-10550.pdf
2021/2021-27114.pdf
2000/00-31313.pdf
2021/2021-14127.pdf
2021/2021-11792.pdf
2018/2018-04860.pdf
2015/2015-15495.pdf
2016/2016-12579.pdf
2007/07-2027.pdf
2004/04-13123.pdf
2006/06-9993.pdf
2023/2023-24283.pdf
2008/E8-17940.pdf
2003/03-5343.pdf
2022/2022-11810.pdf
2021/2021-10460.pdf
2022/2022-22531.pdf
2018/2018-17068.pdf
2013/2013-13523.pdf
2017/2017-28160.pdf
2022/2022-28474.pdf
2019/2019-06548.pdf
2021/2021-28313.pdf
2016/2016-22962.pdf
2023/2023-16570.pdf
2009/E9-31418.pdf
2003/03-7736.pdf
2013/2013-12157.pdf
2007/07-4907.pdf
Download Summary Table
Show 10 entries
Find in Any Column:
document
topic
theta
All
All
All
1999/99-14633.pdf
1
0.9997


http://127.0.0.1:3944/#topic=0&lambda=1&term=
Text Explorer
Documents
Term Use
Sentiment
Topic Model
Word Cloud
Help
LDAvis Introduction
Cleaning
Lemmatize
Stemming
Remove Punctuation
Remove Numbers
Remove Stop Words
Remove Custom Stop Words
Custom Stop Words (single words, comma or newline separated)
Load
Save
ngrams
Minimum Frequency to Include
2
100
2 12 22 32 42 52 62 72 82 92 100
Selected Topic: 0
Previous Topic
Next Topic
Clear Topic
Slide to adjust relevance metric:(2)
λ = 1
0.0 0.2 0.4
Intertopic Distance Map (via multidimensional scaling)
PC2
PC1
1
2
3
4
5
Marginal topic distribution
2%
5%
Top-5 Most Relevant Terms for Topic 3 (21.1
0 1,000 2,000 3,000
homeland security
presidential document
federal register
po frm
fr fm
Overall term frequency
Estimated term frequency within the selected topic
1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topic
2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert &


http://127.0.0.1:3944/#topic=0&lambda=1&term=
Text Explorer
Documents
Term Use
Sentiment
Topic Model
Word Cloud
Help
Cleaning
Lemmatize
Stemming
Remove Punctuation
Remove Numbers
Remove Stop Words
Remove Custom Stop Words
Custom Stop Words (single words, comma or newline separated)
Load
Save
ngrams
Minimum Frequency to Include
2
100
2 12 22 32 42 52 62 72 82 92 100
ngram Length
Uniform Manifold Approximation and Projection (UMAP)
UMAP 3
1
2
3
4
5
Federal Register / Vol. 69, No. 231 / Thursday, December 2, 2004 / Presidential Documents 70177
Presidential Documents
Executive Order 13364 of November 29, 2004
Modifying the Protection Granted to the Development Fund for Iraq and Certain Property in Which Iraq Has an Interest and Protecting the Central Bank of Iraq
By the authority vested in me as President by the Constitution and the laws of the United States of America, including the International Emergency
2004/04-26686.pdf
2003/03-13412.pdf
2005/05-3385.pdf
2014/2014-12651.pdf
2001/01-2140.pdf
2001/01-13381.pdf
2015/2015-13055.pdf
2003/03-7160.pdf
1999/99-17444.pdf
2003/03-19676.pdf


http://127.0.0.1:3944/#topic=0&lambda=1&term=
Text Explorer
Documents
Term Use
Sentiment
Topic Model
Word Cloud
Help
Cleaning
Lemmatize
Stemming
Remove Punctuation
Remove Numbers
Remove Stop Words
Remove Custom Stop Words
Custom Stop Words (single words, comma or newline separated)
fmt
sfmt
federal
Load
Save
ngrams
Minimum Frequency to Include
2
100
2 12 22 32 42 52 62 72 82 92 100
ngram Length
Black & White
department agency
executive department
applicable law
military judge
secretary treasury
court martial
homeland security
national security
act usc
task force
secretary homeland
consistent applicable
director office
jurisdiction datum
arm force
head agency
white house
national emergency
classify information
january presidential
december presidential
office management
management budget
consultation secretary
secretary defense
convene authority
permit law
assistant president
person property
property property
law enforcement
public law
unite america
treasury consultation
personal jurisdiction
october presidential
leader minority:18

Text Explorer

Documents | Term Use | Sentiment | Topic Model | Word Cloud | Help

About
Session Info
How to Cite

Cleaning

How to Cite

```
To cite R in publications use:



To cite package 'tidyr' in publications use:

  Wickham H, Vaughan D, Girlich M (2024). _tidyr: Tidy Messy Data_. R package version 1.3.1,
  <https://CRAN.R-project.org/package=tidyr>.

A BibTeX entry for LaTeX users is

  @Manual{,
    title = {tidyr: Tidy Messy Data},
    author = {Hadley Wickham and Davis Vaughan and Maximilian Girlich},
    year = {2024},
    note = {R package version 1.3.1},
    url = {https://CRAN.R-project.org/package=tidyr},
  }


----------------------------------------


To cite package 'tidylda' in publications use:

  Jones T (2024). _tidylda: Latent Dirichlet Allocation Using 'tidyverse' Conventions_. R package version 0.0.5,
  <https://CRAN.R-project.org/package=tidylda>.

A BibTeX entry for LaTeX users is

  @Manual{,
    title = {tidylda: Latent Dirichlet Allocation Using 'tidyverse' Conventions},
    author = {Tommy Jones},
    year = {2024},
    note = {R package version 0.0.5},
    url = {https://CRAN.R-project.org/package=tidylda},
  }
```

Packages

- shiny
- shinydashboard
- shinybusy
- shinyFiles
- cachem
- DT

I/O

- digest
- readtext
- textdata
- pdftools

Parallel Processing

- foreach
- doParallel
- progressr

Analysis

- tidytext
- textstem
- stringr
- dplyr

- quanteda
- sentimentr

- tidyr
- tidylda
- Matrix

- proxy

- uwot
- geometry
- rnndescent
- FNN

Visualization

- plotly
- LDAvis
- ggplot2
- wordcloud2

Thank you

William Doane
wdoane@ida.org