

All the News That's Fit to Mine: Text Mining and Topic Modeling Newspaper Usage in Scholarly Literature

Mary Feeney, Librarian for Social Sciences & News Research
Niamh Wallace, Librarian for Social Sciences & English Literature
Jeff Oliver, Data Science Specialist



University of Arizona Libraries

Scholars' use of newspapers: past research

Who uses newspapers?

History, Communication Studies, Journalism,
Political Science, Sociology, Arts, Medical,
Archaeology, Geography, Genealogy etc., etc.

Scholars' use of newspapers

But what do they write about? And does this differ depending on newspapers used?

Possible ways to discover this: text mining and topic modeling

Text mining:

Turning text into data for
computational analysis



Search

Search:

Data For Research
beta

Free text mining tool for the
JSTOR archives

dfr.jstor.org



Search

Search: Anywhere in document ▾

Data For Research

beta

Selected: 1,039

Sort By: Relevance ▾

« Prev 1 2 3 ... 104 Next »



Selection Criteria

✕ Clear All

Article Type:

✕ Research article

any:

✕ "Chicago Defender" or "Chicago Daily Defender"

Content Type:

✕ Journal

Narrow results by:

+ Year of Publication

- Content Type

Journal (1,039)

export

- Article Type

Research article (1,039)

export

+ Key terms

+ Journal

+ Publisher

+ Author

+ Language

1. God Didn't Curse Me When He Made Me Black

Christopher Manning

Journal of the Illinois State Historical Society (1998-), Vol. 102, No. 1 (Spring, 2009), pp. 28-72

Subjects: African American Studies, Political Science, Law, American Studies

Key terms: defender, congressional, civil, congress, congressman, amendment, access, representative, party, democratic, veteran, colleague, leadership, debate, segregationist, office, democrat, legislation, politician, picken, session, desegregation, speech, haygood, segregation

More Info

2. Blowing the Trumpet: The "Chicago Defender" and Black Migration during World War I

James R. Grossman

Illinois Historical Journal, Vol. 78, No. 2 (Summer, 1985), pp. 82-96

Subjects: African American Studies

Key terms: defender, migration, southerner, south, newspaper, reader, letter, migrant, league, lynching, mississippi, black, movement, unite, railroad, circulation, exodus, employment, racial, offer, ottley, advertisement, folder, opportunity, themself

CiteRank: 0.0412

More Info

3. The Negro Press in Illinois

Metz T. P. Lochar

Journal of the Illinois State Historical Society (1908-1984), Vol. 56, No. 3, Emancipation Centennial Issue (Autumn, 1963), pp. 570-591

Subjects: African American Studies

Key terms: negro, defender, abbot, chicago, newspaper, white, press, protest, racial, searchlight, black, migrant, discrimination, ottley, south, fight, paper, movement, equality, editorial, headline, american, people, problem, story

More Info

4. Music, Gender, and "Uplift" in the "Chicago Defender", 1927-1937

Lawrence Schenbeck

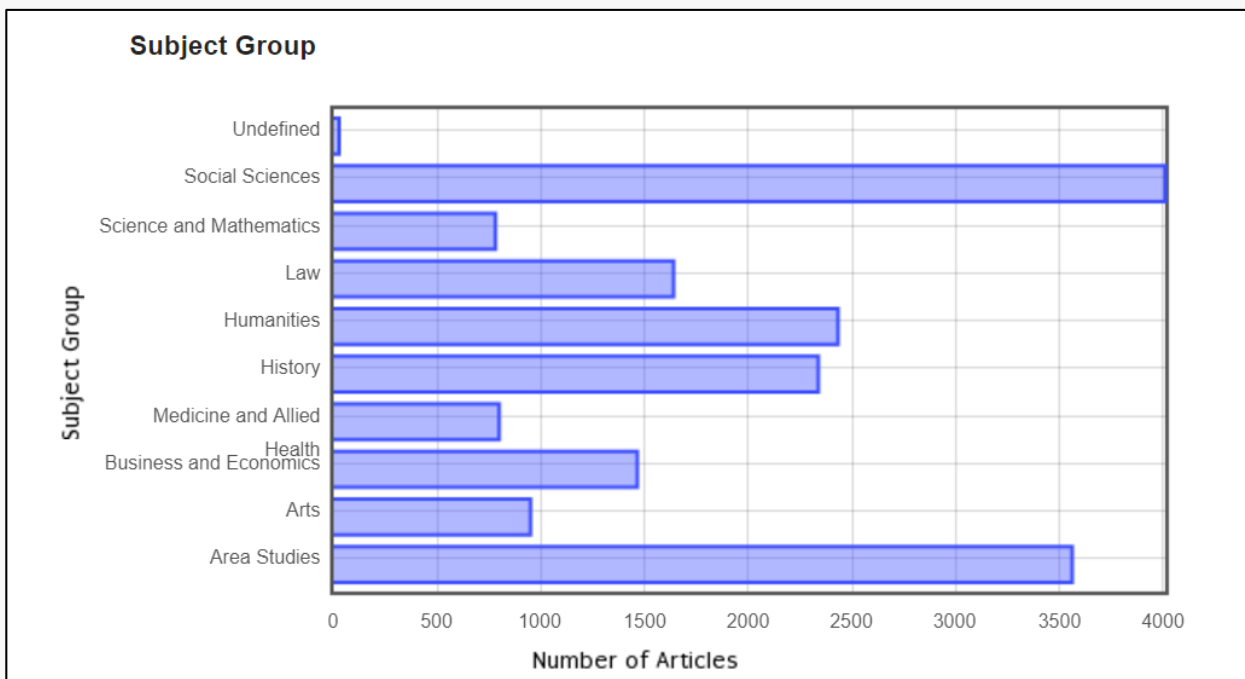
The Musical Quarterly, Vol. 81, No. 3 (Autumn, 1997), pp. 344-370

Subjects: Music African American Studies American Studies

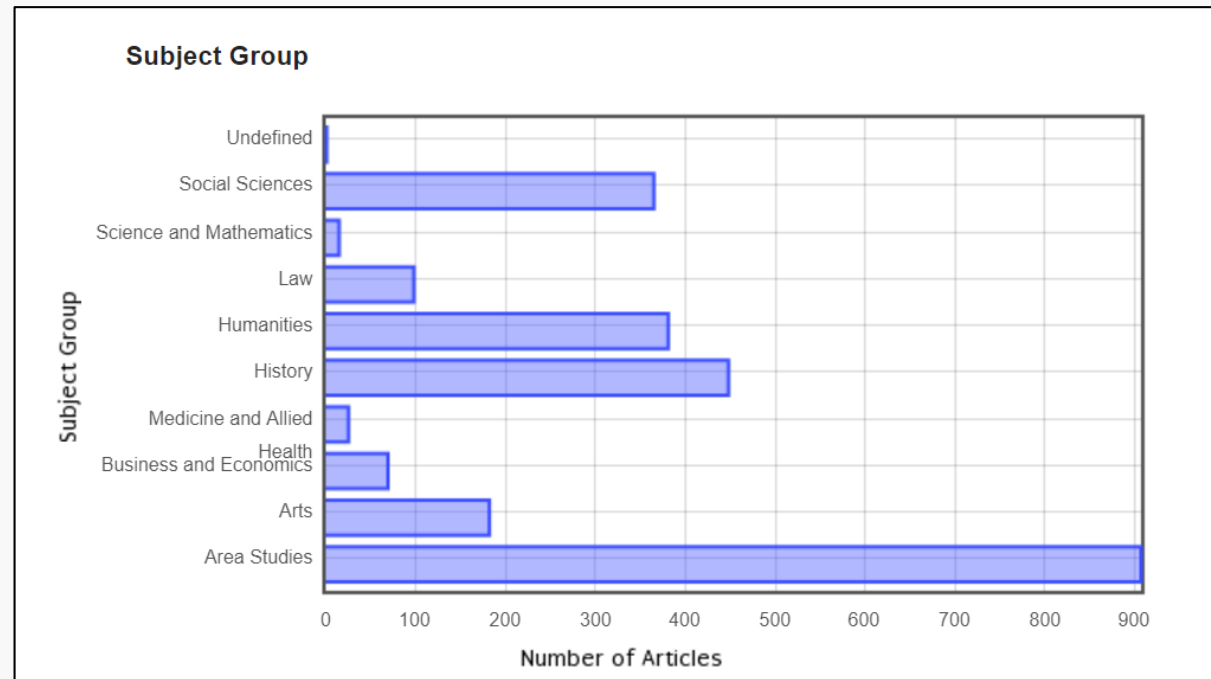
Usage of newspapers in scholarly journals

Newspaper	# of search results in JSTOR DfR
New York Times	99,766
Washington Post	25,477
Chicago Tribune	7,852
Chicago Defender	1,039
Arizona Daily Star	585
El Tucsonense	8
El Democrata Fronterizo	5

Chicago Tribune



Chicago Defender



Key terms of articles using Chicago papers

Chicago Defender

african american black
chicago civil colored community
defender harlem labor **naacp**
negro newspaper people racial
school south state **white** woman
worker

Chicago Tribune

american black **chicago**
child **court** government
newspaper party people percent
political president **public**
republican school **state**
student **tribune** unite white
woman

El Tucsonense



Search

Search: Anywhere in document ▾

Selected: 8

Sort By: Relevance ▾

« Prev 1 Next »

Selection Criteria [✕ Clear All](#)

any:

[✕ "El Tucsonense"](#)

Article Type:

[✕ Research article](#)

Narrow results by:

[Year of Publication](#)

[Content Type](#)

Journal (8)

[export](#)

[Article Type](#)

Research article (8)

[export](#)

- MADRES Y ESPOSAS: Tucson's Spanish-American Mothers and Wives Association**
Julie A. Campbell
The Journal of Arizona History, Vol. 31, No. 2 (Summer, 1990), pp. 161-182
Subjects: Latin American Studies
Key terms: mexican, tucson, woman, ronquillo, association, club, chatter, fierro, mother, war, wife, r tucsonenses, bond, tucsonans, carmen, august, orona, sodality
[More Info](#)
- Spanish-Language Newspapers and Periodicals Published in the United States**
Marjorie C. Johnston
Hispania, Vol. 34, No. 1 (Feb., 1951), pp. 85-87
Subjects: Latin American Studies
Key terms: monthly, weekly, idiom, revista, prensa, informant, nueva, daily, student, ingenieria, idior mundo, petroleo, indicador, panamerican, material, sanitarium, defensor
[More Info](#)
- De factores de inestabilidad nacional a elementos de consolidación del Estado posrevolucion:**
Fernando Saúl Alanis Enciso

[akers](#) [alguno](#) [americano](#) [april](#) [artist](#)
[association](#) [august](#) [barrio](#) [bond](#) [border](#)
[borderland](#) [butterfield](#) [daily](#) [member](#)
[mexican](#) [mexicano](#) [mundo](#)
[people](#) [prensa](#) [tucson](#)
[weekly](#)

El Demócrata Fronterizo



Search

Search: Anywhere in document ▾

Selected: 5

Sort By: Relevance ▾

« Prev 1 Next »

Selection Criteria

✕ Clear All

any:

✕ "El Demócrata Fronterizo"

Article Type:

✕ Research article

Narrow results by:

▣ Year of Publication

▣ Content Type

Journal (5)

export

▣ Article Type

Research article (5)

export

- Deconstructing "La Raza": Identifying the "Gente Decente" of Laredo, 1904-1911**
Elliott Young
The Southwestern Historical Quarterly, Vol. 98, No. 2 (Oct., 1994), pp. 227-259
Subjects: Latin American Studies, Feminist & Women's Studies
Key terms: laredo, decente, mexican, fronterizo, gente, bruni, woman, worker, democ deconstruct, labor, cristiana, nicasio, racial, newspaper, mexicanist, union
More Info
- SARA ESTELA RAMIREZ: SEMBRADORA**
INÉS HERNANDEZ, Sara Estela Ramirez
Legacy, Vol. 6, No. 1, Western Women Writers (Spring 1989), pp. 13-26
Subjects: Latin American Studies, Feminist & Women's Studies
Key terms: magon, aurora, woman, fronterizo, marfa, mexican, flores, democrata, cro chicana, chicano, newspaper, essay, siempre, revolution
More Info
- "VÁMONOS PA' MÉXICO" LA COMUNIDAD MEXICANA EN ESTADOS UNIDOS Y L**
1917-1918
Fernando Saúl Alvarado Enciso

angel anglo astrea attack aurora
battle birthday border chicana
chicano democrata elite
fronterizo laredo
mexican mexico mujer
newspaper racial texas
woman



Please complete this form to request a dataset. The request will be sent to the JSTOR Support Team, who will reply to confirm your request and give you a time estimate for delivering it.

Please note:

- These requests may include only metadata and ngrams for journals and pamphlets on JSTOR. For full-text dataset requests, please [contact JSTOR Support](#) for more information.
- Returning Data for Research users: The self-service option is currently offline, but you may obtain similar datasets via this request form. Some data options have changed, keywords and .csv format files are not currently available.

[About the Data](#)

JSTOR incorporates metadata and OCR text from many different sources. We work to mitigate any inconsistencies, but you may come across issues with datasets, such as duplicate OCR, missing OCR, or incomplete metadata. We are unable to assist with the clean-up of datasets due to OCR errors or missing data; however, if you suspect an error in the processing of the dataset, please contact us.

Project Description

Enter brief description*

2. I would like to request the following data:

All datasets include item-level metadata

Per-Document N-Grams

- Word counts
- Bigrams
- Trigrams

3. What is the scope of the data to be provided?

JSTOR Search URL, if one has been constructed

The dataset will be limited to these publication dates:

The dataset will be limited to these journals (please separate titles with commas):

Topic modeling:

“a computational, probabilistic technique to uncover categories and discover patterns in and among texts”

From Mining the Dispatch (<http://dsl.richmond.edu/dispatch/>)

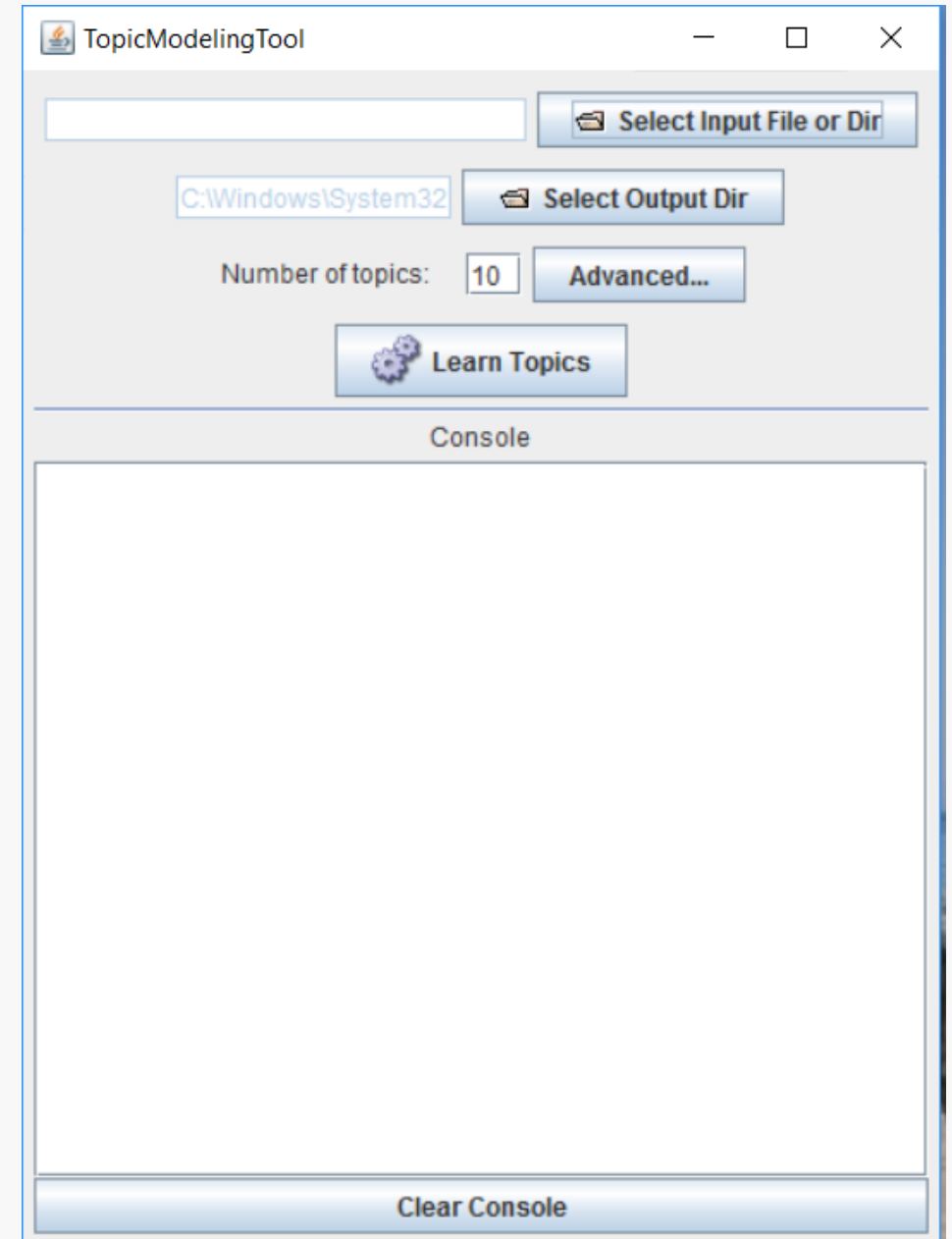
“a method for finding and tracing clusters of words (called ‘topics’ in shorthand) in large bodies of texts”

<http://miriamposner.com/blog/very-basic-strategies-for-interpreting-results-from-the-topic-modeling-tool/>

Tools for Topic Modeling



<https://code.google.com/archive/p/topic-modeling-tool/downloads>



Process

- Ran MALLET with:
 - 500 iterations
 - 10 topic words in each topic
 - 20 topics
 - Used their default stop word list

The screenshot shows the TopicModelingTool interface. At the top, there are two text boxes for input and output directories, each with a "Select Input File or Dir" button. The input directory is "C:\Users\mfeeney\Documents\USTOR DfR topic modeling\Chicago Defender\Chicago Defender data" and the output directory is "C:\Users\mfeeney\Documents\USTOR DfR topic modeling\Chicago Defender". Below these is a "Number of topics:" field set to "20" and an "Advanced..." button. A "Learn Topics" button with a gear icon is centered below. The bottom section is a "Console" window displaying the following text:

```
Chose Input Directory: C:\Users\mfeeney\Documents\USTOR DfR topic modeling\Chicago Defender\Chicago Defender data
Chose Output Directory: C:\Users\mfeeney\Documents\USTOR DfR topic modeling\Chicago Defender
Importing and Training...this may take a few minutes depending on collection size.Importing from: C:\Users\mfeeney\Documents\UST
[-remove-stopwords, true, --preserve-case, false, --input, C:\Users\mfeeney\Documents\USTOR DfR topic modeling\Chicago Defend
Labels =
  C:\Users\mfeeney\Documents\USTOR DfR topic modeling\Chicago Defender\Chicago Defender data
[-num-iterations, 500, --num-top-words, 11, --doc-topics-threshold, 0.05, --input, C:\Users\mfeeney\Documents\USTOR DfR topic mo
Data loaded.
Coded LDA: 20 topics, 5 topic bits, 11111 topic mask
max tokens: 2662070
total tokens: 37284975
<10> LL/token: -10.80303
<20> LL/token: -9.99769
<30> LL/token: -9.85659
<40> LL/token: -9.79985

0          2.5          women black white american african race men social racial class
1          2.5          baseball sport team sports players robinson game black american louis
2          2.5          black american white african americans war york race negro racial
3          2.5          black negro white blacks south negroes chicago southern city race
4          2.5          lynching war army soldiers military die der mob men und
5          2.5          school university students education schools history college black american african
6          2.5          chicago city black community public housing urban local racial white
7          2.5          mixed citation id type publication label ref fn journal xlink
8          2.5          source group year person string fpage ref surname italic type
```

Topics of articles that used *Chicago Defender*

List of Topics

1. women white black men american woman gender race social class
2. baseball black sport team american sports players robinson game league
3. black white american racial african americans race york negro rights
4. negro negroes black white blacks colored south race washington men
5. war army soldiers die der military japanese und veterans soldier
6. school university students education schools college history american african black
7. chicago black city community public housing urban african local american
8. mixed citation id publication type label fn ref journal york
9. source group year person string fpage ref surname italic type
10. reading news newspapers media newspaper test children subjects group research
11. cultural culture social identity sense race community power historical ways
12. black film american hughes white york films african harlem micheaux
13. lynching police law state violence case crime white trial men
14. africa african american world du bois garvey states war robeson
15. black migration blacks social american population south migrants southern labor
16. music jazz blues band black musical musicians dance york american
17. de la en le les di des il del se
18. american art jewish jews religious african chicago religion culture church
19. state court law rights public board committee federal case act
20. man people time back good life work family day don

Topics of articles that used *Chicago Tribune*

List of Topics

1. american social history world political religious life society people church
2. art building chicago museum design city work york buildings water
3. data research analysis table model number journal results information study
4. school education students schools university reading college teachers research children
5. war american united states foreign world government military international german
6. business labor company market economic tax trade industry public land
7. women culture cultural woman world sexual identity body language image
8. president political party state committee congress house public republican election
9. york american pp book review ed books literary library john
10. black chicago white women american social city african race racial
11. de la en el los del di las se por
12. police law crime criminal trial case cases death court justice
13. news music film press york media newspaper newspapers times american
14. de la le les des en du une dans est
15. court law note supra state id states act rights federal
16. chicago illinois lincoln state war history county john july men
17. health medical child children care research family human women social
18. time man people mr day made men good life years
19. international policy political government military states nuclear http security state
20. der die und von den des zu das im mit

Topics of articles that used *Chicago Defender*

List of Topics

1. women white black men american woman gender race social class
2. baseball black sport team american sports players robinson game league
3. black white american racial african americans race york negro rights
4. negro negroes black white blacks colored south race washington men
5. war army soldiers die der military japanese und veterans soldier

Topics of articles that used *Chicago Tribune*

List of Topics

1. american social history world political religious life society people church
2. art building chicago museum design city work york buildings water
3. data research analysis table model number journal results information study
4. school education students schools university reading college teachers research children
5. war american united states foreign world government military international german

Labeling and interpreting topics

Chicago Defender - Topics	Topic label
women white black men american woman gender race social class	Gender and race
baseball black sport team american sports players robinson game league	African Americans in sports
black white american racial african americans race york negro rights	
negro negroes black white blacks colored south race washington men	
war army soldiers die der military japanese und veterans soldier	

Labeling and interpreting topics

Chicago Tribune - Topics	Topic label?
american social history world political religious life society people church	Church and state
art building chicago museum design city work york buildings water	Chicago arts and design
data research analysis table model number journal results information study	
school education students schools university reading college teachers research children	
war american united states foreign world government military international german	

What does this tell us?

Initial thoughts:

- Researchers tend to utilize major national newspapers to the exclusion of a wider and more diverse range of newspapers.
- Scholars using *Chicago Defender* vs. *Chicago Tribune* tend to address different research questions.
- Would this hold to be true in articles that use other newspaper titles?
- Do the topics change over time based on trends in scholarly research?

Mary Feeney, mfeeney@email.arizona.edu

Niamh Wallace, niamhw@email.arizona.edu

Jeff Oliver, jcoliver@email.arizona.edu