

TOPIC MODELING ENSLAVEMENT NARRATIVES

Jessica Shinker

Topic Title (Tentative)	Weight	Topic Words
Memory	1.34194	time man good day told people place house made thought found money asked white night great make home friends days
Human Nature	0.88397	time made character state good subject means present life power free liberty nature human case persons fact act mind give
Human Relationships	0.75881	children master slave family heart poor freedom free long life mother years wife found friends great felt hope brought mind
Enslavement - Place	0.64524	master day time night house place man large great country work water made field road ground slaves corn cotton men
Enslavement - Family	0.51701	master man slaves back told slave mother time free put work night children sold day white made men wife house
Religion - Spirituality	0.47054	god lord man great thy love soul christ world heaven jesus life spirit meeting power death thou thee earth heart
Women, Embodiment	0.45361	young face mistress eyes room hand heart love mother poor girl child woman miss long make thing door looked moment
The State - White Supremacy	0.39909	colored men man race years war white states great united people state south made negro city life country government president
Religion - Institution	0.21879	church rev conference baptist years pastor bishop colored school year members work methodist city preach john churches episcopal people time
Enslavement - State, Structure	0.20867	slaves slave slavery negro states men south law man state poor slaveholders negroes years virginia southern county laws masters carolina
International Slave Trade	0.20519	great africa country negro england coloured america african english people men man negroes captain board king london british years native
Education	0.19708	people school negro work race white south washington time man education years life tuskegee schools students country men great large
The State - Violence, Control	0.18366	court john state case law man judge states house boston city street burns slave lewis united madison county day act
Religion - Community	0.1818	lord god mrs home church work meeting people brother jesus sister thought mother time pray good night told mission heart
Ranching, Cowboys	0.15782	men time years prison life man smith home indian cattle long william massa started back brown boys indians horse horses
Enslavement - Abolition	0.13491	slavery slave douglass man anti-slavery people slaves states men american master country great free baltimore part power brown liberty work
Crops, Food	0.10834	farm bond cotton scott water put sugar uncle make corn sweet add land carver work boss cut seed crop negro
Haiti	0.07168	toussaint french general men france hayti domingo l'ouverture island blacks west put country cape government saint war army cadet order
Dialect	0.06689	dat miss aunt don't peter ter dey yer i'll dis bout git den didn't sally wid dar i'd ole dice
Individuals	0.024	mrs matthias folger sojourner isabella truth pierson lincoln letter friend good dear sancho woman god n't york family kingdom spirit

Pros and Cons

Topic modelling can be problematic. MALLET's inner workings are difficult to understand, and it's often misused by scholars. There are also Black Feminist critiques of using big data methods to analyze Black voices - when misinterpreted or misapplied, results can be misleading or outright harmful.

One of topic modeling's biggest strengths is as a brainstorming mechanism. Notice the topic labeled "Ranching, Cowboys"? That's a topic I may not have noticed by myself. It's especially interesting for connecting enslavement, ranching, and Natives. MALLET also made an intriguing distinction between religious spirituality, institutions, and communities. Other topics - Haiti, Food, Women & Embodiment - could make for more cool research projects.

Trial and Error...

One of my hypotheses was that modeling these texts as individual grafs rather than full texts might make for a more satisfying model. But in the end, there wasn't a clear advantage to splitting texts.

To the right are the topic names I assigned to my graf-by-graf model and the Python script I wrote to split the texts into paragraphs.

This experiment didn't work out, but it proved to me how useful programming skills can be for digital historians.

Topic Title (Tentative)
Memories
Daily Life
Embodied Emotion
Slavery - Philosophy
Religion - Spirituality
Family
Body
Education
The State - Washington
**Sojourner Truth, Narrative
Capitalism
Riverboats
Abolition
Enslavers
Food
Religion - Institution
Intern'l Slave Trade
The State - Courts
Dialect
Haiti

Introduction

I used the program MALLET to topic model DocSouth's collection of slave narratives. MALLET identifies words that often occur near one another and notes how frequently these "topics" are mentioned across all texts. The researcher then tentatively assigns a name to each topic. This can be hard - topics are often vague.

Topic modeling can reveal patterns in texts that are otherwise missed.

```
1 # code to split an entire text (memoir, biography, etc.)
2
3 import re
4 import os
5
6 # change working directory (and check to make sure it works)
7 malletDir = 'c:\\mallet\\hist-data\\texts'
8 os.chdir(malletDir)
9 print("Current working directory:", malletDir)
10
11 # create a new folder for the split texts
12 savepath = malletDir + '\\textsgrafs'
13 os.mkdir(savepath)
14
15 # cycle through each text in the folder
16 for text in os.listdir(malletDir):
17
18     # open text file, read it into a variable,
19     # remove bracketed info, and write each paragraph into
20     # with open(text, 'r', encoding="utf8") as file:
21
22         # read file name into variable and check
23         title = text
24         print(title)
25
26         #read file into variable, close the file, and
27         var = file.read()
28         file.close()
29         pass
30
31     # remove bracketed text (useless image information)
32     modified_var = re.sub(r"\\[\\.\\?\\]\\", "", var)
33
34     # split text into paragraphs
35     splat=modified_var.split("\\n")
36
37     # for each individual paragraph, name it and write it
38     for number, paragraph in enumerate(splat, 1):
```