

Reviews in Digital Humanities

Review: On the Books: Jim Crow and Algorithms of Resistance

Ann Marie Blackmon¹, Carolina Collins¹

¹University of Texas at Austin

Published on: Feb 08, 2021

License: [Creative Commons Attribution 4.0 International License \(CC-BY 4.0\)](https://creativecommons.org/licenses/by/4.0/)

Project

On the Books: Jim Crow and Algorithms of Resistance

Project Directors

[Amanda Henley](#), University of North Carolina

[Matthew Jansen](#), University of North Carolina

Project URL

<https://unc-libraries-data.github.io/OnTheBooks/>

Project Reviewers

[Ann Marie Blackmon](#), University of Texas at Austin

[Carolina Collins](#), University of Texas at Austin

Project Overview**Amanda Henley and Matthew Jansen**

On the Books: Jim Crow and Algorithms of Resistance is a Collections as Data and machine learning project inspired by a K-12 teacher who contacted a librarian in search of a comprehensive listing of all North Carolina Jim Crow laws.

The project created text corpora of North Carolina session laws and used machine learning techniques to discover Jim Crow laws passed between Reconstruction and the Civil Rights Movement (1866-1967). A website provides searchable access to the Jim Crow laws and contextualizes them with an essay and a collection of K-12 learning resources. This project relied on the Python programming language and open source software, and a GitHub site hosts the scripts written for the project. Documented examples from the workflow are provided in Jupyter notebooks. Our workflow is detailed in [the white paper](#), and generally follows these simplified steps: acquisition, adjustment and manipulation of digitized images, OCR, corpus segmentation, analysis of the corpora using supervised machine learning, XML generation, and corpus creation.

The [project team, consultants, and collaborators](#) consist of librarians, scholars, and information professionals providing a wide range of expertise, including: text analysis, coding, visualization, digital scholarship, metadata, legal information, web design, software development, project management, K-12 education, OCR, history, and African American studies. The first phase of the project (10/19–08/20) was part of *Collections as Data Part to Whole* (funded by Andrew W. Mellon Foundation). Phase two will conclude May 2020 and is funded through the Association of Research Libraries.

We envision multiple audiences for this project: information professionals interested in creating collections as data, legal scholars interested in North Carolina laws, anyone interested in learning more about Jim Crow laws, and K-12 educators interested in teaching about Jim Crow laws. The project has been presented widely to librarians, digital humanists, and K-12 teachers. We are hopeful that the promotion of the project will engage a broad audience. The initial products of this project were released August 31, 2020. In 17 days, the Jim Crow text corpus was downloaded 33 times, the white paper was downloaded 43 times, and we were informed that an undergraduate student is using On the Books products for their undergraduate honor's thesis. An essay about the project was [published in *Black Perspectives*](#) by team member William Sturkey.

On a larger philosophical level, On the Books acknowledges the implicit bias of algorithms and aims to use them to purposely expose racism. Safiya Noble's *Algorithms of Oppression* (NYU Press, 2018) has revealed how algorithms are implicitly biased by the people who code them, arguing that Google's search algorithms reinforce racism. Can we, as information professionals, counter this bias? If we acknowledge there are algorithms of oppression, could there also be algorithms of resistance? On the Books successfully developed algorithmic approaches to discover racist laws, but we are also clear about the limits of the algorithmic approach: the identification of Jim Crow laws can be subjective, and the true force of Jim Crow existed and persists far beyond algorithmic detection.

Project Review

Ann Marie Blackmon and Carolina Collins

On the Books: Jim Crow and Algorithms of Resistance identifies and offers access to Jim Crow era laws passed in North Carolina from 1866 to 1967 that discriminated against both African Americans and Indigenous people. Seeking to answer the question, "Can text mining and machine learning identify racist language in legal documents?" On the Books successfully illustrates how laws and codes written after the Civil War but before the Civil Rights Movement contain racist rhetoric and word choice. This rhetoric detrimentally influenced the lives of Black people during this hundred-year period in American history. It also explores how optical character recognition (OCR), algorithms, and machine learning technologies used to analyze the Jim Crow law corpus express bias and racism in their operation. A cross-section of UNC Libraries employees with diverse backgrounds involving data analysis, data visualization, content development, text analysis, and statistics and disciplinary scholars collaborated to ensure the project both served its various audiences and met existing standards.

On the Books uses Python and open source software to identify and transform digitized images of laws passed by the North Carolina legislature over a hundred-year period that have been made available by the Internet Archive. Algorithms run against the Internet Archive generated two plain-text corpuses:

1) all North Carolina Session Laws from 1866-1967 and 2) Jim Crow laws enacted by North Carolina. As acknowledged by project members, shortcomings of the identifying algorithm ultimately preclude a completely comprehensive survey of Jim Crow laws, though 905 Jim Crow laws are represented in the Jim Crow-specific text corpus. The algorithm also encountered issues with the inclusion of some false positives in the corpus. From there, the corpus was transformed through optical character recognition to turn digitized images into machine-readable text. On the Books then systematically mined each corpus using topic modeling and supervised classification to determine racist wording in official laws cast by North Carolina and to increase the searchability of the text corpora. Clarification on where human intervention was required beyond the supervised classification would assist those interested in deploying the same algorithm and workflow. Digital humanities researchers will find that publication of the project's Python tutorials through Jupiter notebooks will be particularly useful. Published via the project's GitHub repository, researchers can easily fork On the Books for their own research or classroom use.

Beyond the GitHub repository, the project provides access to a white paper on the project, a timeline, primary and secondary source materials, and lessons that allow teachers to relay information about the Jim Crow laws to students. Contextual essays also make the content accessible to researchers and others interested in exploring Jim Crow laws passed in North Carolina. Some of the materials, like the lessons, timeline, and "The Laws in Context" page, serve to contextualize the Jim Crow laws by discussing their historical precedent and impact on African American and Native communities. Of special note is On the Books' collaboration with the UNC Department of History and Carolina K-12's director and manager, which resulted in outstanding resources for K-12 curricula. As such, On the Books engages audiences of varying ages and backgrounds to identify and make accessible heretofore disregarded and silenced histories through technological means.