The Vanderbilt University CENTER FOR DIGITAL humanities

African American Arts Coverage in the Black Press

PI: Rebecca VanDiver (History of Art and Architecture), Mellon Faculty Fellow in the Digital Humanities

Collaborators: Sarah Swanz (formerly VU library), Ali Yaqoob (Data Sciences)

Abstract

"African American Arts Coverage in the Black Press" is a text and data analysis project focused on understanding how the Black Press (periodicals, newspapers, magazines, etc.) covered African American visual artistic production (painting, sculpture, photography, etc) during the twentieth century.

Our principal research questions are aimed at understanding how the Black press was reporting on developments in the African American visual arts scene: who, what, when.

Which visual artists were being mentioned and in what context?
What types of visual art received coverage?
Are there chronological trends in arts coverage?
Does this information match up with the scholarship on twentieth century African American artistic production

Charles White, *The Awakening*, 1961

Data Sets Used

A major hurdle this project faced was gaining access to the data since it all falls under copyright. The well-timed receipt of a PROQUEST TDM Studio Research Grant enabled the team to utilized copyrighted materials in the ProQuest created environment.

ProQuest Historical Newspapers: Black Periodicals Database. Created chronological slices of data (1919-1940 and 1960-1970) to test methods on materials when African American visual production was high. Using ProQuest TDM Studio environment limited the data set size to under 2,000,000 documents.

- *Afro-American*
- *Chicago Defender*

Methods

In order to answer the research questions we had posed, our aim was to look through the text within the historical newspapers and find the articles where African American Artists were being mentioned.

There were two major limitations in this project:

1. The articles were read into the system through Optical Character Recognition (OCR) and therefor the quality of the text being parsed was questionable.

2. The workspace environment was heavily limited due to limited connection to the internet which meant all major text processing tools (Spacy, NLTK, etc.) had to be manually downloaded in order to be used.

In order to assess the quality of OCR, we searched through the articles and created a readability metric. This metric was essentially a measure of the proportion of recognizable English words that existed with each article.
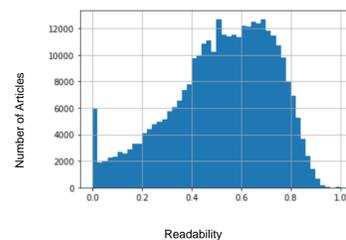
Figure 1: Proportions of articles with different variabilities in a single publication

Search Algorithm

We created a list of artist names, the variations they could have in how they are mentioned in articles, the periods they were mentioned and their gender.

The algorithm was to go through each article, look at the year it was published and search for names of artists that we expected to be there based on when they started being active.

We employed two major methods to look for these names:

1. A **strict string match** which counted exact mentions of artists in the articles.
2. A more relaxed ("fuzzy") search which accounted for the type mistake we could potentially see due to the poor OCR quality.

We tried both to find that the first method was the most reasonable and resulted in the fewest noisy articles that were being captured.

Results and Analysis

With the search algorithm we looked into (among other questions):

1. How many artist mentions were happening each year
2. Who was being mentioned the most. Who was being mentioned the least.
3. The disparity in mentions based on the gender of the artist.
4. What were the co-occurrences of mentions of artists (who was being mentioned together)

Initial results indicate a correlation between the spikes in visual arts mention and major periods in African American art history (e.g., Harlem Renaissance of the 1920s)

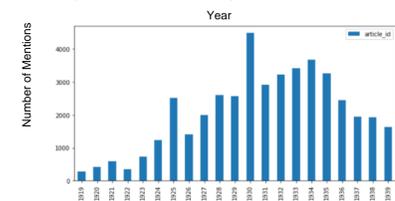Below are some figures representing what we looked into.

Figure 2: Number of mentions over the years. (Using the fuzzy method)

	Artist 1	Artist 2	Artist 3	Artist 4
Artist 1	-	20	1	0
Artist 2	20	-	1	5
Artist 3	1	1	-	0
Artist 4	0	5	0	-

Table 1: Simplified representation of the artist co-occurrences.

Next Steps

Some next steps involve incorporation of both the search mechanisms and the readability score to develop a unified approach. This unified approach would also take into account the number of art related terms being mentioned in order to increase the quality of matches we are getting (i.e. is the article being captured really an art related article. In the end we should get a score associated with how likely it is that a given article has a given mention in it and whether that mention was truly something we can trust.

Acknowledgements