GRAPHING YALE’S THEATRICAL HISTORY

Catherine DeRose
Yale University Library Digital Humanities Laboratory

Alex O’Keefe
Robert B. Haas Family Arts Library
Why Ensemble@Yale?

THE SUBJECT: Yale Theatre History

- DRA 37 Archival Collection
- Department / School of Drama - 1925
- Yale Repertory Theatre - 1966
- 100+ world premieres
- Famous alumni
  - Meryl Streep, Wendy Wasserstein, Angela Bassett, William Ivey Long, Liev Schreiber, Sigourney Weaver, Michael Yeargan, Frances McDormand, John Guare, Lupita Nyong’o...

DESIRED OUTCOMES:

- Increase access to minimally-processed archives
- Generate database of Yale theater history
- Answer reference questions
Project Background

Inspired by NYPL Labs’ Ensemble project

- Limits of OCR
- Importance of human judgment
- Semantic relationships
- Hard for computers to see
- Easy for humans to intuit
Project Background

**TASKS:**
- Digitize programs
- Determine organization and presentation
- Adapt Scribe with customized workflows for theatrical data
  - Mark
  - Transcribe

**LESSONS LEARNED:**
- Usability testing before going live
- Fixing errors/preventing junk data
- How to decide when marking or transcribing is completed?
Launched May 2017

Help create a database of Yale theater history
Click a button below to get started:

MARK
TRANScribe

Browse 90 years of programs to see what's available to mark or transcribe:
Demo - Marking
Demo - Transcribing
Dealing with the Data

Steps for Completed Programs

● Initial Cleaning [OpenRefine]
● Removing Duplicates [Excel]
● Program Check [student worker / Excel]
● Normalization to Improve Searching
  ○ Names and Roles
    ■ Roles example: shift from ‘scenery’ to ‘set design’
  ○ Choosing name authorities
    ■ AusStage, IBDB, Playbill Vault, Library of Congress, etc.
● Final Check
## Standardizing Data with OpenRefine

### 2430 rows

<table>
<thead>
<tr>
<th>All</th>
<th>Set Key</th>
<th>Title</th>
<th>Year</th>
<th>Opening</th>
<th>Closing</th>
<th>Person [Original]</th>
<th>Person [Normal]</th>
<th>Role [Original]</th>
<th>Role [Normal]</th>
<th>Role [Stage]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>42307-00-001-0004-01</td>
<td>The Singing Minute</td>
<td>1927</td>
<td>11/6/27</td>
<td>11/22/27</td>
<td>Homer Abeggen</td>
<td>Abeggen, Homer</td>
<td>Mr. Snell</td>
<td>Mr. Snell</td>
<td>Cast</td>
</tr>
<tr>
<td>2</td>
<td>42307-00-001-0005-02</td>
<td>What's A Fixer For?</td>
<td>1928</td>
<td>5/1/28</td>
<td>5/5/28</td>
<td>Homer Abeggen</td>
<td>Abeggen, Homer</td>
<td>Stage Carpenter</td>
<td>Scenic Artist</td>
<td>staff</td>
</tr>
<tr>
<td>3</td>
<td>42307-00-001-0005-03</td>
<td>All People</td>
<td>1928</td>
<td>5/1/28</td>
<td>5/5/28</td>
<td>Homer Abeggen</td>
<td>Abeggen, Homer</td>
<td>Stage Carpenter</td>
<td>Scenic Artist</td>
<td>staff</td>
</tr>
<tr>
<td>4</td>
<td>42307-00-001-0005-02</td>
<td>The Crab</td>
<td>1928</td>
<td>5/1/28</td>
<td>5/5/28</td>
<td>Homer Abeggen</td>
<td>Abeggen, Homer</td>
<td>Stage Carpenter</td>
<td>Scenic Artist</td>
<td>staff</td>
</tr>
<tr>
<td>5</td>
<td>42307-00-001-0005-02</td>
<td>Minnie Field</td>
<td>1928</td>
<td>5/1/28</td>
<td>5/5/28</td>
<td>Homer Abeggen</td>
<td>Abeggen, Homer</td>
<td>Stage Carpenter</td>
<td>Scenic Artist</td>
<td>staff</td>
</tr>
<tr>
<td>6</td>
<td>42307-00-001-0005-03</td>
<td>Brand</td>
<td>1928</td>
<td>5/28/28</td>
<td>5/29/28</td>
<td>Homer Abeggen</td>
<td>Abeggen, Homer</td>
<td>Properties Master / Mistress</td>
<td>Staff</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>42307-00-001-0005-02</td>
<td>It's Morning</td>
<td>1940</td>
<td>2/5/40</td>
<td>2/7/40</td>
<td>Herbert Abrahams</td>
<td>Abrahms, Herbert</td>
<td>Slaves</td>
<td>Slaves</td>
<td>cast</td>
</tr>
<tr>
<td>8</td>
<td>42307-00-001-0003-01</td>
<td>Machiavelly</td>
<td>1927</td>
<td>4/30/27</td>
<td>4/30/27</td>
<td>Theodore C. Achilles</td>
<td>Theodore C.</td>
<td>First Soldier</td>
<td>First Soldier</td>
<td>Cast</td>
</tr>
<tr>
<td>9</td>
<td>42307-00-001-0003-01</td>
<td>Machiavelly</td>
<td>1927</td>
<td>4/30/27</td>
<td>4/30/27</td>
<td>Theodore C. Achilles</td>
<td>Theodore C.</td>
<td>Lepers</td>
<td>Lepers</td>
<td>Cast</td>
</tr>
<tr>
<td>10</td>
<td>42307-00-001-0003-01</td>
<td>Machiavelly</td>
<td>1927</td>
<td>4/30/27</td>
<td>4/30/27</td>
<td>Theodore C. Achilles</td>
<td>Theodore C.</td>
<td>Soldiers</td>
<td>Soldiers</td>
<td>Cast</td>
</tr>
<tr>
<td>11</td>
<td>42307-00-001-0004-04</td>
<td>Finished</td>
<td>1928</td>
<td>2/13/28</td>
<td>2/14/28</td>
<td>T.C. Achilles</td>
<td>Achilles, Theodore C.</td>
<td>People on Train</td>
<td>People on Train</td>
<td>Cast</td>
</tr>
<tr>
<td>12</td>
<td>42307-00-001-0004-04</td>
<td>Pablo</td>
<td>1928</td>
<td>2/29/28</td>
<td>3/1/28</td>
<td>Theodore C. Achilles</td>
<td>Theodore C.</td>
<td>Shamans</td>
<td>Shamans</td>
<td>Cast</td>
</tr>
<tr>
<td>13</td>
<td>42307-00-001-0004-07</td>
<td>Finished</td>
<td>1928</td>
<td>3/15/28</td>
<td>3/16/28</td>
<td>T.C. Achilles</td>
<td>Achilles, Theodore C.</td>
<td>People on Train</td>
<td>People on Train</td>
<td>Cast</td>
</tr>
<tr>
<td>14</td>
<td>42307-00-001-0005-03</td>
<td>Brand</td>
<td>1928</td>
<td>5/28/28</td>
<td>5/29/28</td>
<td>T.C. Achilles</td>
<td>Achilles, Theodore C.</td>
<td>People of the Hamlet</td>
<td>People of the Hamlet</td>
<td>Cast</td>
</tr>
<tr>
<td>15</td>
<td>42307-00-001-0002-05</td>
<td>All The King's Men</td>
<td>1949</td>
<td>12/14/49</td>
<td>12/16/49</td>
<td>Robert Ackart</td>
<td>Ackart, Robert</td>
<td>Stage Crew</td>
<td>Stage Crew</td>
<td>Staff</td>
</tr>
<tr>
<td>16</td>
<td>42307-00-001-0002-06</td>
<td>Welcome Home</td>
<td>1929</td>
<td>5/6/29</td>
<td>5/11/29</td>
<td>Tony Acri</td>
<td>Acri, Tony</td>
<td>Coriolanus Sons</td>
<td>Coriolanus Sons</td>
<td>cast</td>
</tr>
<tr>
<td>17</td>
<td>42307-00-001-0002-04</td>
<td>Lazy</td>
<td>1927</td>
<td>1/31/27</td>
<td>2/27/27</td>
<td>Thomas J. Ahearn</td>
<td>Ahearn, Thomas J.</td>
<td>Moses Newton</td>
<td>Moses Newton</td>
<td>Cast</td>
</tr>
<tr>
<td>18</td>
<td>42307-00-001-0002-04</td>
<td>Lazy</td>
<td>1927</td>
<td>1/22/27</td>
<td>2/22/27</td>
<td>Thomas J. Ahearn</td>
<td>Ahearn, Thomas J.</td>
<td>Moses Newton</td>
<td>Moses Newton</td>
<td>Cast</td>
</tr>
<tr>
<td>19</td>
<td>42307-00-001-0002-07</td>
<td>Twelve Before Three</td>
<td>1927</td>
<td>3/16/27</td>
<td>3/19/27</td>
<td>Thomas J. Ahearn</td>
<td>Ahearn, Thomas J.</td>
<td>Playwright</td>
<td>Playwright</td>
<td>playwright</td>
</tr>
<tr>
<td>20</td>
<td>42307-00-001-0002-07</td>
<td>Machiavelly</td>
<td>1927</td>
<td>4/30/27</td>
<td>4/30/27</td>
<td>Thomas J. Ahearn</td>
<td>Ahearn, Thomas J.</td>
<td>Lepers</td>
<td>Lepers</td>
<td>Cast</td>
</tr>
<tr>
<td>21</td>
<td>42307-00-001-0002-01</td>
<td>Machiavelly</td>
<td>1927</td>
<td>4/30/27</td>
<td>4/30/27</td>
<td>Thomas J. Ahearn</td>
<td>Ahearn, Thomas J.</td>
<td>Soldiers</td>
<td>Soldiers</td>
<td>Cast</td>
</tr>
</tbody>
</table>
## Standardizing Names

### Cluster & Edit column "Person [Original]"

This feature helps you find groups of different cell values that might be alternative representations of the same thing. For example, the two strings "New York" and "new york" are very likely to refer to the same concept and just have capitalization differences, and "Gödel" and "Godel" probably refer to the same person. **Find out more ...**

<table>
<thead>
<tr>
<th>Cluster Size</th>
<th>Row Count</th>
<th>Values in Cluster</th>
<th>Merge?</th>
<th>New Cell Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>8</td>
<td>Herschel V. Williams Jr. (6 rows) Herschel V. Williams Jr. (3 rows)</td>
<td>✓</td>
<td>Herschel V. Williams Jr.</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>Robert MacKay (2 rows) Robert MacKay (1 row)</td>
<td>✓</td>
<td>Robert MacKay</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>Keith Mackaye (4 rows) Keith Mackaye (3 rows)</td>
<td>✓</td>
<td>Keith Mackaye</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>Richard H. Diggs Jr (3 rows) Richard H. Diggs Jr (1 row)</td>
<td>✓</td>
<td>Richard H. Diggs Jr</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>Edwin Bradley Jr (1 row) Edwin Bradley, Jr. (1 row)</td>
<td>✓</td>
<td>Edwin Bradley Jr</td>
</tr>
</tbody>
</table>

### Additional Features

- **Cluster Size**
- **Row Count**
- **Values in Cluster**
- **Merge?**
- **New Cell Value**

### Statistics

- **Rows in Cluster**
- **Average Length of Choices**
- **Length Variance of Choices**
### Processing Data: Creating Edges

```python
import csv
import itertools
from collections import defaultdict

peopleList = []
x = 0

with open("2019-05-08-drama.csv", "r", encoding='UTF-8') as reader:
    writer = csv.writer(output)
    headers = next(reader)
    headers.append('id')
    writer.writerow(headers)

    for row in reader:
        # add unique ID for each person[normalized]
        person = row[6]
        if person not in peopleList:
            peopleList.append(person)
            x += 1
            print(person, x)
        row.extend([str(x)])
        writer.writerow(row)
```

```python
import csv
import itertools
from collections import defaultdict

d = defaultdict(list)
plays = {}

with open("2019-05-08-drama-ids.csv", "r", encoding='UTF-8') as reader:
    writer = csv.writer(output)
    output.writerow('source' + ',' + 'target' + ',' + 'label' + 'id')
    headers = next(reader)

    # populate dictionary
    for row in reader:
        d[row[0]] + '-' + row[1] + '-' + row[2].append(row[10])

    # create combinatorial edges
    for key, value in d.items():
```
Experiments with Networks
Experiments with Networks: Layout

Force Atlas 2

Fruchterman Reingold
Experiments with Networks: Color
Experiments with Networks: Size

Betweenness Centrality

Eigenvector Centrality
Experiments with Networks: Time

Animating connections: 1925 to 1952
Experiments with Networks: Partitions

- Cast
- Directors
- Staff

Modularity Group 2

Modularity Group 0
Finding Individual Networks
Thank You!

Questions?

Catherine DeRose
Digital Humanities Lab Manager
catherine.derosse@yale.edu

Alex O'Keefe
Arts Digital Projects Librarian
alexandra.okeefe@yale.edu

Acknowledgments: Paul Beaudoin | Tess Colwell | Douglas Duhaime | Mauricio Giraldo | Lindsay King | Peter Leonard | Francesca Livermore | Steven Padla | Alexandra Provo | Danielle Reay | Monica Ong Reed | Doug Reside | Ben Vershbow | Maria Zapata

ensemble.yale.edu