Computational Curation of a Digitized Record Series of WWII Japanese-American Internment

William Underwood
Richard Marciano
Sandra Laib

Computational Archival Science Workshop #2
IEEE International Conference on Big Data 2017
December 13, 2017, Westin Copley Plaza, Boston, Massachusetts
Issued by President Franklin Roosevelt on February 19, 1942, this order authorized the evacuation of all persons deemed a threat to national security from the West Coast to relocation centers further inland. In the next 6 months, over 100,000 men, women, and children of Japanese ancestry were moved from Oregon and Washington state to assembly centers. They were then evacuated to and confined in isolated, fenced, and guarded relocation centers, known as internment camps.
The National Archives and Records Administration (NARA) is the repository of the records of this program. Record Group 210, Records of the War Relocation Authority, includes paper records of internal security cases and associated paper index cards for the 10 Relocation Centers. These records have not been released to the public due to access restrictions on some of the records.
Objectives

- Curate the information in these cards by improving their quality (scanning, OCR, text correction, analysis, and extraction) and adding value to the repository of digital information by providing to NARA descriptive metadata supporting access to items in this record series.

- Extract item-level metadata from the Relocation Center index cards in order to supply archivists at the National Archives and Records Administration (NARA) with the information needed to support withdrawal decisions for items in this record series.

- Explore archival analytics approaches through social networking analysis and geospatial processing of the information extracted from the index cards.
The DCIC is Pursuing a CAS Training / Teaching Agenda

There is a need to:

○ create innovative classes that emphasize new modes of collaboration, and interdisciplinary work.

○ blend elements of archival thinking and computational thinking:
  ○ problem solving that uses modeling, decomposition, pattern recognition, abstraction, algorithm design, and scale.

○ develop inter-disciplinary iSchools with faculty from Computer Science, Archival Science, and Data Science.

○ develop extensive hands-on experience working with cyberinfrastructure to carry out archival functions.

WE WELCOME PARTNERSHIPS -- CONTACT US!
Incident Index Cards
Analysis of the Tule Lake Index Cards

Index card to case reports

<table>
<thead>
<tr>
<th>Incident date</th>
<th>Case report ID</th>
<th>Subject</th>
<th>Japanese-American or Japanese internee name</th>
<th>Residence ID</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>11-4-43</td>
<td>A-999</td>
<td>Riot</td>
<td>Amesu, Otsuki 9999-D</td>
<td>One of possible leaders or trouble makers.</td>
<td></td>
</tr>
</tbody>
</table>

Meaning of the information on card

- Incident date,
- the case report ID (A-999),
- the relevant page number in the report (P9),
- the subject of the case report (incidents that are offenses such as a Riot),
- the Japanese-American or Japanese internee name,
- the residence ID in the camp (9999-D) and
- a remarks section
Analysis: Other Information on Index Cards

○ names of administrative and staff members of the center
○ document identifiers (C-816)
○ document types (letter, memo)
○ names of relocation center facilities (hospital, stockade)
○ job titles (chief cook, #35 Block Manager, secretary)
○ organizations such as young Japanese men’s militant groups (Hokoku Seinen Dan)
○ relocation center organizations (Administrative Police, Agricultural Division),
○ locations (Honolulu, Hawaii, San Francisco, Calif.)
○ time and time periods (2:45 pm, 36 days)
○ Remarks: relations such as membership in militant organizations and actions/events (arrested, interviewed by, sentenced to time period, and released).
Styles of Index Cards

- Index card to Case Reports
  - With Other IDs (Family number (F#8504), Individual number (8504-A))
  - With Multiple Names
  - With Other Cases (Accidents, Transfers)
- Index card to Case Report using Keywords (Name, Subject, Location: Remarks)
- Index card to Registration Arrests (Name, Arrested, Sentenced, Released)
- Index card to Documents (Memo, Teletype, Letter)
- Continuation of Remarks – Remarks on back of Index Card
The above and [redacted] were put in project jail for military marching, blowing of bugles, display of Japanese emblems. Occurred in the colony.
Text Annotation using the ANNIE Pipeline

<table>
<thead>
<tr>
<th>Name</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Document Reset PR</td>
<td>Document Reset PR</td>
</tr>
<tr>
<td>ANNIE English Tokeniser</td>
<td>ANNIE English Tokeniser</td>
</tr>
<tr>
<td>ANNIE Gazetteeer</td>
<td>ANNIE Gazetteeer</td>
</tr>
<tr>
<td>Tule Gazetteer</td>
<td>ANNIE Gazetteeer</td>
</tr>
<tr>
<td>ANNIE Sentence Splitter</td>
<td>ANNIE Sentence Splitter</td>
</tr>
<tr>
<td>ANNIE POS Tagger</td>
<td>ANNIE POS Tagger</td>
</tr>
<tr>
<td>ANNIE NE Transducer</td>
<td>ANNIE NE Transducer</td>
</tr>
<tr>
<td>Tule NE Transducer</td>
<td>ANNIE NE Transducer</td>
</tr>
<tr>
<td>ANNIE OrthoMatcher</td>
<td>ANNIE OrthoMatcher</td>
</tr>
</tbody>
</table>
### Wordlists Specific to the Index Cards

<table>
<thead>
<tr>
<th>List name</th>
<th>Major</th>
<th>Minor</th>
<th>Language</th>
<th>Annotation type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Japanese_Militant_Organizations.lst</td>
<td>organization</td>
<td>militant</td>
<td></td>
<td>Lookup</td>
</tr>
<tr>
<td>Japanese_female_given_names.lst</td>
<td>person_first</td>
<td>female</td>
<td></td>
<td>Lookup</td>
</tr>
<tr>
<td>WAR_cities.lst</td>
<td>location</td>
<td>city</td>
<td></td>
<td>Lookup</td>
</tr>
<tr>
<td>WAR_location_facilities.lst</td>
<td>facility</td>
<td>building</td>
<td></td>
<td>Lookup</td>
</tr>
<tr>
<td>japanese_male_given_names.lst</td>
<td>person_first</td>
<td>male</td>
<td></td>
<td>Lookup</td>
</tr>
<tr>
<td>japanese_names_initialcaps.lst</td>
<td>person_full</td>
<td></td>
<td></td>
<td>Lookup</td>
</tr>
<tr>
<td>offenses.lst</td>
<td>offense</td>
<td></td>
<td></td>
<td>Lookup</td>
</tr>
<tr>
<td>tule_lake_job_titles.lst</td>
<td>jobtitle</td>
<td>tulelake</td>
<td></td>
<td>Lookup</td>
</tr>
<tr>
<td>tule_lake_organizations.lst</td>
<td>organization</td>
<td>tulelake</td>
<td></td>
<td>Lookup</td>
</tr>
</tbody>
</table>
Color-Coded Annotations and Annotation List for a Sample Index Card

11-4-43 A-999 P9 Riot
Amane, Ohashi
9999-D
One of possible leaders or trouble makers.

<table>
<thead>
<tr>
<th>Type</th>
<th>Set</th>
<th>Start</th>
<th>End</th>
<th>Id</th>
<th>Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Date</td>
<td>0</td>
<td>7</td>
<td>101</td>
<td></td>
<td>{kind=date, rule=DateNumDash, ruleFinal=DateOnlyFinal}</td>
</tr>
<tr>
<td>CaseReportId</td>
<td>17</td>
<td>22</td>
<td>106</td>
<td></td>
<td>{rule=CaseReportId_Page}</td>
</tr>
<tr>
<td>CaseReportPage</td>
<td>23</td>
<td>25</td>
<td>107</td>
<td></td>
<td>{rule=CaseReportId_Page}</td>
</tr>
<tr>
<td>Offense</td>
<td>40</td>
<td>44</td>
<td>108</td>
<td></td>
<td>{rule=Offenses}</td>
</tr>
<tr>
<td>Person</td>
<td>47</td>
<td>60</td>
<td>109</td>
<td></td>
<td>{rule=RuleReverseNamePersonNoSplit}</td>
</tr>
<tr>
<td>Residence</td>
<td>62</td>
<td>68</td>
<td>110</td>
<td></td>
<td>{rule=ResidenceAddress}</td>
</tr>
</tbody>
</table>
Interpretation of the Annotated Index Cards

(qlf =[

  filename( "Box8-Tule_Lake-999.jpg" ),
  style( "Box8-Tule_Lake-999.jpg",
        "index_card_to_case_reports" ),
  case_no( "Box8-Tule_Lake-999.jpg","A-999" ),
  page_no( "A-999", "9" ),
  date( "A-999", "11-4-43" ),
  subject( "A-999", "Riot" ),
  person_name( "Amane, Ohashi" ),
  gender( "Amane, Ohashi", "male" ),
  indexname( "Amane, Ohashi",
             "Box8-Tule_Lake-999.jpg" )
  residence_no( "Amane, Ohashi", "999-D" )
])
Visual Analytics of the Annotated Index Cards
Questions
Workflow for Automated Review of the OCRed Index Cards for Restrictions on Release to Public

- **Scan, OCR and correct OCR errors** of the "index cards".
- **Annotate** the person’s names, gender, and, if included, the age, family number, and individual number of the persons named on the OCRed incident cards.
- **Interpret the annotated terms** using an Ontology of the concepts on the index cards and export the interpretation.
- **Check the person names and associated information** against the full database of internee information (also part of RG210), which includes person names, birthdates, family numbers and individual numbers.
- **Recommend index cards for withdrawal** with person names and information matching information in the internee database of juveniles (under 18) and the release to the public of the other cards.