

Tule Lake Computational History

Background

During World War II, 110,000 and 120,000 Japanese Americans in the United States were relocated and incarcerated in camps. Throughout imprisonment, American police and camp security guards recorded incidents involving the camp prisoners in small cards. In addition to falsifying events, the cards often contained racist connotations. The DCIC current has 10,000 digital images of the cards and 109,000 data points discussing the people, places, and events.

Objectives

The Tule Lake project aims to develop a graph database to model the internment camp instance cards. Furthermore, the project aims to develop Natural Language Processing flows for big archival record data extraction using the Densho WRA records. This would result in training a machine to learn and extract Japanese names.

These objectives will be developed through three, independent phases:

1. Data Modeling and Graph Database Brainstorming
2. Natural Language Processing Work Flows developed by augmenting WRA registry
3. Building Graph Database

Theme

Citizen Internment

Stakeholders

Bill Underwood – Principle Affiliate DCIC Research Scientist

Marcus Most – Office of Innovation, National Archives

Japanese Americans- Community

Data

Archival Images - 10,000 out of 25,000 images

Excel Sheet with 109,000 detainee data

JPEG's/PDF's

TXT files

Final Product:

Graph Database

NLP Work Flow for Computational Archives

Skills:

Programming

- Natural Language Processing for Data Extraction
- Python
- Javascript

Machine Learning

Data Analytics / Data Management / Data Modeling

Human Computer Interaction

Database building

Software:

Gate Installations

NEO4J

R

WordPress

Professional Development:

Papers

Conferences

Historical Society Networking

Additional Resources

<http://ddr.densho.org/names/>

<https://catalog.archives.gov/id/1264228>

<http://local.ads.nwsourc.com/ads/FlippingBook/2015/Q2/Densho/html/>