**Uncover marginalized narratives of Japanese American Incarceration with text mining**

**University Of North Texas**

**INFO 5731**

**Haihua Chen & Roopesh Maganti**

# 

# ABSTRCT

**Srujan Bandla**

University of North Texas Denton, Texas, USA

# Shubhika kanakaboina

University of North Texas Denton, Texas, USA

# Megan collins

University of North Texas Denton, Texas, USA

# Dakshayani

# Rayini

University of North Texas Denton, Texas, USA

# Sadia

# Mahbub

University of North Texas Denton, Texas, USA

During World War II, the United States interned around 120,000 Americans of Japanese ancestry. 904 recorded interviews regarding "Japanese American incarceration experience from people who lived it" are the majority of the oral histories in the Densho Digital Repository's collection. The demographic data for 904 narrators—including their race, age, and location—were gathered through online scraping techniques for this study. With 78.82% of all narrators being Nisei, she represents the largest proportion of Japanese Americans. When the incarceration took place in 1942, there were 279 narrators (or 39.35%) who were minors (born between 1926 and 1942), with more than half of them (58.78%) being teenagers (age between 12 and 17). The west coast, specifically Washington State, Oregon State, or California State, is home to the bulk of narrators (82.09% of the total). This study also develops a framework for incarceration oral history, in which the historical texts related to incarceration can be processed manually, automatically, or both ways. There are seven categories in the annotation scheme: (1) the pre-war context of the imprisonment; (2) the government's choice to expel ethnic Japanese; (3) living after expulsion and while imprisoned; (4) military duty; (5) the return of ethnic Japanese after WWII; (6) judicial challenges; and (7) the redress movement.

# KEYWORDS

Japanese American incarceration; Digital archives; Natural language processing; Annotation scheme.

# GIT HUB LINK

# INTRODUCTION

President of the United States Franklin D. Roosevelt signed Executive Order 9066 ten weeks after the Japanese attack on Pearl Harbor on December 7, 1941, imprisoning some 120,000 persons of Japanese descent as a result. Citizens of the United States made up two-thirds of those compelled to live in military camps. Following a government study in the 1980s, it was determined that Americans of Japanese descent had experienced a serious injustice. A nonprofit group called Densho was established to record oral history testimony from Japanese Americans in order to preserve the history of the Japanese American Incarceration as people's recollections of it fade (Densho, n.d. a). To "increase civic engagement and educate about and advance democratic ideas," Densho created the Densho Digital Repository (DDR)( (Beckman & Froh, 2018). The "Japanese American incarceration experience from individuals who lived it" is covered in 904 recorded interviews that are part of the oral history archive that DDR has put together (Densho, n.d. b). For convenience, every interview has been completely transcribed and divided into subtopics of watching. The collection also includes demographic data about the narrators. More researchers are using computer techniques to further their study and hunt for hidden patterns in massive datasets, such as text mining, network analysis, and data visualization.

A semi-structured starting dataset with a high-quality and domain-specific corpus is needed for computation analysis. A clear annotation strategy (Chen, 2022). An automatic or manual process known as corpus annotation procedure of incorporating interpretive data into a corpus of texts to improve it (Hovy & Lavid, 2010). An codebook known as the annotation scheme outlines the annotation categories and procedures Lavid, 2010). Chen (2022) suggests the following as a workable method for computationally processing and mining the texts: select an annotation strategy, annotate a little amount of early, high-quality data, and then add data using a large-scale unlabeled data using a semi-supervised learning technique. This study makes two contributions using the oral history collection from the DDR as our corpus: (1) This study initially investigates the backgrounds of those who have experienced the past and shared their narratives in order to elucidate the suppressed history of Japanese American Incarceration. The demographic data of the narrators, including their nationality, race, age, and locality, is specifically examined in this study. (2) Based on the fine-grained annotation scheme we suggest, the historical writings about the Japanese internment can be handled either manually or mechanically

**RESEARCH QUESTIONS:**

People should not be judged solely based on their looks or their ancestry. Search Information without allowing a person's perspective to override underlying assumptions. That made our country. It's important to avoid keeping things quiet and personal. This is the story It became unspeakable for a long time. It has an important place in our collective psyche and tradition. These are the issues that will be looked at

* How did Japanese internment camps affect society?
* How did internment affect Japanese American families?
* What was the purpose of internment camps in the US?
* How did Japanese Americans respond to incarceration?
* What was daily life like in the incarceration camps?
* What were some of the other perspectives involved?

# LITERATURE REVIEW:

In 1941, there were over 100,000 people of Japanese heritage living in the mainland United States, most of them at that time lived on the coast of the Pacific Ocean. Most of the participants are born and raised in the United States and, therefore, were full citizens (Commissarchik, Sen, & Velez, 2020). As a result, Roosevelt's anti-Japanese reaction after the attack on Pearl Harbor The administration changed its stance toward Japanese Americans, both citizens and foreigners. The government violated the constitutional rights of Japanese Americans fixed justification for this policy was based on the belief that national security was at stake. Japanese Americans have been under surveillance by the Office of Naval Intelligence and the FBI ever since A total of more than 3,000 suspected saboteurs were arrested by these two a agencies and the Army's G-2 intelligence unit after the 1930 Pearl Harbor attack (Karolin & Aden, 2021). National security agencies saw no reason to worry about the remaining Japanese American population. On the other hand, people were skeptical. A 1942 report claimed that Japanese Americans leaked information to Japan before Pearl Harbor contributed to anti-Japanese hysteria in the United States. Many Pacific Coast residents feared that Japanese Americans living in their area might help the Japanese military launch an invasion of their area (Camp-Whittaker & Clark, 2019). Despite public pressure, officials were wary of detaining Japanese Americans, especially citizens, without clear justification. At the time, neither Attorney General Francis Biddle nor Secretary of War Henry Stimson thought removal was a good idea or even legal under the law. Even at the highest levels of the military, military officials maintained that this policy was necessary to ensure public safety on the Pacific coast. It was a nightmare of displacement and uncertainty for Japanese Americans Implicit Japanese Americans had one week to issue an exclusion order for registration with the authorities, collect their belongings and report to the nearest assembly center (Takezawa, 2019). Many migrants had to give up their possessions within days. Thus, Homeowners and business owners need to sell their property quickly and cheaply, often thousands of people were waiting in stables, animal shops and in the open-air transfer them to designated detention camps when converted from assembly center racetracks and fairgrounds. The unjust detention of Japanese Americans is a psychological wound and fixed an unfair reputation that lingered long after the war ended. From the sad reference point, what these Americans describe changes in their surroundings, personal lives, and aspirations." An American Progress Reporting Center. Personnel, ethnic, historical, and cultural traumas are all present, and each has significant influence on the other (Arellano-Bover, 2022). The time, Intergenerational trauma formed from incarceration, personal and racial trauma in the post-war period history and contemporary trauma emerge. For many Japanese Americans, the sudden uprooting fear and uncertainty of being deported or imprisoned without a crime, without a crime their future destroyed their preconceived notions of who they were. When it comes to racial profiling, Japanese and Italian Americans are singled out, but this was not the case for Germans and Italian Americans. Many anti-Asian laws were enacted because the Japanese had a belong history of anti-Japanese prejudice before World War II Unreliable and unable to assimilate into the country's culture. For the most part, Japanese Americans advocate the pursuit of justice. But those who were in society who feared that the group's “making waves” would rekindle old animosities (Arellano-Bover, 2022). The pain and suffering of Japanese Americans in receiving the money was not insignificant compensation. It also reflects ongoing tensions with the Japanese American Citizens League Necessary sections of appropriate routes for remediation. Psychology’s long history of promoting social justice is equally relevant to incarceration. Many Japanese Americans' experiences with imprisonment have prompted psychologists to think about the impact of race trauma on individual distinctions, family and multigenerational processes, and coordinated responses. Having a psychology "deeply stranded in history and values" and attending to the calmness around memories associated with massive social and political disturbance. On the one hand, the trauma of incarceration serves as a call to action for psychological studies into the mechanisms that underlie racial and ethnic prejudices. American citizens' vengeance toward Japanese Americans for a violation the Japanese Committed, Japanese Americans were exposed to harsh treatment during the Second World War. As a result of the Japanese attack on Pearl Harbor, the American people were gripped with paranoia and resentment (Chan, 2019). The U.S. government's concern of a Japanese invasion on the West Coast caused the deportation of Japanese Americans to internment camps (Nagata, Kim & Wu, 2022). In perspective, the internment of Japanese Americans was a disgrace and unnecessary. But the expulsion of this troop was part of a larger plot to defeat an aggressive Japan, especially in the setting of a surprise attack on a defenseless people. Many Japanese Americans battled in Europe against Axis Powers, the removal of innocent Japanese Americans from their residential and commercial buildings without Compensating them is disgusting. People worried about a possible assault from Japan on the US the west coast coast, after Pearl Harbor focused their attention on the Nisei population in America. As a result, they spent the rest of the Pacific War in detention camps. Specific communities along the western seaboard of the United States had notice boards that read "Civilian Exclusion Order," which was the title of the order. These bulletins served as notice to all residents of Japanese heritage that they would be vacating their residences (Yamasaki, 2020). The location of they’re the final destination was unknown because the announcement did not specify where they were going. Individuals who were summoned to the central unit were asked to provide the names and ages of their entire families (Collins, 2018). They were then told to gather with the rest of their family at a designated time and location. On a vacation to a foreign land, the Japanese Americans could only bring hand-carried luggage. Numbers have substituted the names of all of the Families. When these numbers were substituted for their names, they lost their sense of self. Families were given less than two weeks to either lease their property or sell all they owned. Japanese Americans and Japanese immigrants alike are experiencing fixed despair and humiliation at the hands of the government. Many Japanese American farmers possessed the land, but they were forced to give it all up due to the eviction. Overcrowding and substandard living circumstances plagued the Japanese Americans held in concentration camps in the United States. War Relocation Authority records show that Japanese Americans were held in tarpaper-covered barracks with no plumbing or internet access after the war, the administrative organization. Because these camps were constructed in a short period during the summer of 1942 by civilian contractors using military barracks designs, the buildings were inadequately suited for crowded living conditions (Stanutz, 2018). Coal was in short supply, and folks slept under as many blankets as they could get their hands on. Nisei, or Japanese Americans born in the United States, were the only ones considered for leadership positions in the camps. The term "Issei" was used to refer to those who were born in Japan and grew up there. Japanese Americans of the Sansei generation were referred to as the "third generation." While the government gave financial support for their children's education, they were forced to remain silent. The only way for concentration camp inmates to leave was to serve in the United States Military, which was the sole choice open to them. The WRA gave all school-age residents of the repositioning facility educative opportunities. fixed courses had already been created, and tutors had been employed to assist the state departments of education in implementing them. At the relocation facility, professional training was provided to improve communication with the grownups. This instruction was intended for migrants who could contribute more effectively to agriculture or business outside of the evacuation centers. At the relocation center, consumer firms were set up to allow internees to buy goods and amenities not offered by the World Relocation authority (WRA). At the relocation centers, religious practices were observed. Over half of the evacuees represented Christian beliefs. After the government forcibly removed Japanese Americans from their homes and interned them in internment camps, some Japanese Americans began to doubt their allegiance to the United States. Despite this, those Japanese Americans have stayed staunchly loyal to the United States of America. In order for evacuees to leave a relocating center, they had to take a job or set up a permanent home. The first step is to study the evacuee's conduct in the relocation site and other data provided by the World Refugee Organization (UN refugee agency). The migrant was denied the right to leave the country if there was any evidence that they posed a threat. As part of the second condition, officials, or residents should have a reasonable basis for believing they were being transferred for a specific purpose. Only those who had a place to go, or some helps were allowed to leave the building. The final criterion was that survivors must notify WRA of any changes in employment or residence. Ethnocidal minorities' experiences are characterized by a long history of racial prejudice that led to their marginalization and imprisonment. A growing body of evidence suggests that most people unwittingly reinforce social disparities by categorizing others as either "us" or "them" based on their group membership. This inadvertent categorization is a contributing factor to acts of racism and discrimination. To deal with these issues, we must continue our efforts to comprehend these procedures and recognize the circumstances for eliminating bias. Finally, his arrest underscores the need to research cross-group partnerships and community activism in response to racial trauma. When the movement for imprisonment redress was just getting started in the 1960s, Japanese Americans joined forces with African American activists to confront civil rights issues. Today, they focus on aiding Muslim and Arab American communities who face continuous hostility and mistrust because of the hazards and consequences of unjust incarceration.

**DATA PREPARATION AND ANALYSIS:**

We retrieve 904 oral history transcripts from the DDR website using web scraping techniques. We utilize the Python Beautiful Soup module to extract the data due to the variety of data pieces seen on each web page, such as paragraph and list elements. Then, we iterate, fetch all interview transcripts, and store them in JSON format for each narrator's web page. Using the spacy phrase matching technique, we additionally extract the narrator's race, birth year, and birthplace from their bio page. The text data that was obtained by dividing the tokens is then examined. Not every bio contains all the necessary demographic data. We extract the race information from 846 narrators, the year of birth information from 709 narrators, and the geographic location information from 832 narrators.

**FINDINGS:**

847 narrators self-identify as Americans, whereas 3 of the 850 narrators self-identify as Japanese.One of the 847 American narrators leaves out her race. 49 narrators are white, making up 5.79% of all American narrators, while 797 (94.21%) are Asian Americans. 796 of the 797 Asian Americans who narrate the stories are Japanese Americans, and the bulk of them (670, or 78.82% of all the narrators, and 84.17% of Japanese American narrators) are Nisei (second-generation Japanese Americans). 709 narrators provide their birth year. The Japanese American Internment occurred in 1942, when 389 (54.87%) of the narrators were adults. 13.4% of the population was adult (aged 26 to 50, born between 1893 and 1925), while 41.47% were young adults (18 to 25). In contrast, 279 (39.35%) of the narrators in 1942 were under the age of 18, with more than half of them (58.78%) being teens (age between 12 and 17). 832 narrators provide information on their native towns. 42 (5%) of them are from Hawaii, while 683 (82.09%) of them are from the west coast, specifically Washington, Oregon, or California.

We develop a seven-category annotation scheme based on the content of two venerable historical works on the internment of Japanese Americans. The books are A Tragedy of Democracy: Japanese Confinement in North America by Greg Robinson and Wendy Ng's Japanese American Internment During World War II: a History and Reference Guide. Both Ng and Robinson demonstrate the history of Japanese American internment from the following perspectives: (1) Pre-war background to the incarceration; (2) Life after removal and during the incarceration; (3) Military services; (4) Legal challenges; (5) Redress movement. We include these five categories in our scheme. Ng's book also introduces the returning of ethnic Japanese from camps after WWII, and we include it as a category of our scheme. Robinson’s unique perspective on how the government decided to remove ethnic Japanese is also included in our scheme. Figure 1 presents a more detailed explanation of each category with definitions. We also demonstrate a transcript example of each category, which will help future researchers to understand how each category can be annotated using the oral history transcripts.

Diagram

Description automatically generated

**Figure 1: Annotation scheme for the history of Japanese American incarceration**

## **MULTI\_LEVEL CLASSIFICATION:**

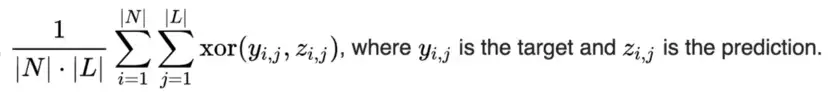
Multi-label classification was developed as a result of research into the text categorization issue, where one document may simultaneously belong to numerous predetermined subjects. Textual data classification with many labels is a significant issue.

In multi-label classification, the aim is to predict the label sets of unseen examples by evaluating training instances with known label sets. The training set is made up of instances, each of which is associated with a set of labels. In contrast to multi-class issues, where the classes are mutually exclusive, multi-label problems describe each label as a new classification task that is somehow related.

**METHODOLOGY:**

## **HAMMING LOSS:**

The simplest definition of Hamming-Loss is the percentage of labels that are mistakenly predicted or the proportion of inaccurate labels to all labels.



**Figure 2: Hamming loss**

## **BINARY RELEVANCE:**

One single-label binary classifier is trained for each class in this instance's ensemble. Each classifier forecasts either a class's membership or non-membership. The result for multiple labels is taken to be the union of all anticipated classes. This method is well-liked because it is simple to use, but it also disregards potential relationships between class labels.

In other words, the binary relevance technique creates n new data sets from the images, one for each label, then trains single-label classifiers on each new set of data if there are n labels. The "binary" in "binary relevance" refers to the fact that a classifier may provide a yes/no response to the question "does it include trees?" Despite being straightforward, this method does not function effectively when there are relationships between the labels.

Graphical user interface, text, application

Description automatically generatedGraphical user interface, text, application, email

Description automatically generatedGraphical user interface, text, application

Description automatically generated

Text

Description automatically generatedGraphical user interface, text, application

Description automatically generated

**Figure 3: Binary Relevance model**

## **CLASSIFIER CHAINS:**

The construction of a chain of binary classifiers C0, C1,..., Cn involves using the predictions of each classifier Ci, where j i. This allows the technique, also known as classifier chains (CC), to account for label correlations.

This method requires the same number of classifiers overall as classes, but the classifiers require more extensive training.

The classification problem with three categories (C1, C2, and C3) linked in that sequence is presented in the example that follows.

Diagram

Description automatically generated

**Figure 4: Classifier chains**

**Graphical user interface, text, application

Description automatically generatedGraphical user interface, text, application

Description automatically generated**

**Graphical user interface, text, application, email

Description automatically generated**

**Graphical user interface, text, application, email

Description automatically generated**

**Graphical user interface, text, application, email

Description automatically generated**

**Figure 5: Classifier chains model**

## **LABEL POWERSET MODEL:**

This method does consider potential relationships between class labels. Because it treats each label in the training set's power set of labels as a separate label, this method is more often known as the label-powerset method.

This method has a significant computational complexity and requires worst-case (2|C|) classifiers.

However, as the number of classes rises, the number of unique label combinations has the potential to multiply exponentially. This quickly causes a combinatorial explosion, which makes computation impossible.

Graphical user interface, text, application

Description automatically generated

Graphical user interface, text, application

Description automatically generatedGraphical user interface, text, application

Description automatically generatedGraphical user interface, text, application, email

Description automatically generatedGraphical user interface, text, application

Description automatically generated

**Figure 6: Power Set model**

**CONCLUSION:**

In this work, the Japanese American internment experience is examined through an oral history collection of 904 interviews with narrators. Particularly, we examine 904 narrators' demographic data. In the meantime, we develop an annotation technique that can be used to analyze the historical literature about incarceration either manually or automatically. Building machine learning models to automatically process the textual oral history will be part of future effort, as well as annotating the interviews according to the plan.

**AUTHOR CONTRIBUTIONS:**

* Srujan Bandla – worked on annotations, Code and ppt
* Shubhika kanakaboina – Worked on annotations, Code and report
* Sadia Mahbub – Worked on Annotations, Code and ppt
* Dakshayani Rayini- Worked on annotations, Code and report
* Megan Collins- Worked on annotations.

**REFERENCES:**

Han, Y., Maganti, R., Chen, J., & Chen, H. (2022). Uncover Marginalized Narratives of Japanese American Incarceration: An Annotation Scheme for Natural Language Processing and Data Analytics. Proceedings of the Association for Information Science and Technology, 59(1), 698–700. <https://doi.org/10.1002/pra2.695>

Beckman, S & Froh, G. (2018). Densho Digital Repository: Preserving Community Memory. Retrieved from (<http://downloads.alcts.ala.org/ce/20180426_Densho_Digital_Repository_Slides.pdf>)

Chen, H. H. (2022). Data quality evaluation and improvement for machine learning. [Doctoral dissertation, North Texas University].

Densho. (n.d. a). About Densho. Retrieved from (<https://densho.org/about-densho/>).

Densho. (n.d. b). Narrators. Retrieved from (<https://ddr.densho.org/narrators/>).

Hovy, E., & Lavid, J. (2010). Towards a ‘science’of corpus annotation: a new methodological challenge for corpus linguistics. International journal of translation, 22(1), 13-36.

Yamasaki, K. (2020). Fish for Jimmy: Inspired by One Family's Experience in a Japanese American Internment Camp. Holiday House.

Murray, A. Y. (2022). Historical memories of the Japanese American internment and the struggle for redress. Stanford University Press.

Komisarchik, M., Sen, M., & Velez, Y. (2020). The Political Consequences of Ethnically Targeted Incarceration: Evidence from Japanese-American Internment During WWII.

Karolin, A. J., & Aden, R. C. (2021). Identifying home: a narrative of Japanese American internment. Communication Quarterly, 69(2), 192-213.

Kamp‐Whittaker, A., & Clark, B. J. (2019). 10 Social Networks and the Development of Neighborhood Identities in Amache, a WWII Japanese American Internment Camp.

Archeological Papers of the American Anthropological Association, 30(1), 148-158.Takezawa, Y. I. (2019). Breaking the silence: Redress and Japanese American ethnicity. Cornell University Press.

Lee, H. C. (2020). Diamonds at the margin of empires:‘baseball complex’in the Japanese

American internment and colonized Taiwan. Cultural Studies, 1-24.

Miyake, J. M. (2021). Japanese American internment experience and the impact on parent-child

relationships. Pepperdine University.

Michel, S. (2019). The Effects of Intergenerational Trauma on Resiliency and Post-traumatic

Growth Among Japanese American Internment Survivors and Their Families (Doctoral

dissertation, Pepperdine University).

<https://www.docsteach.org/documents/document/report-japanese-evacuation-west-coast>

<https://www.docsteach.org/documents/document/exclusion-order>