



CATALOG

[View email in your browser](#)

[Share this email with a friend](#)

New Search Feature: Optical Character Recognition (OCR)

With more than 92 million pages of digitized records available to search in the National Archives Catalog, we are always working on ways to improve search results to better help you find what you're looking for.

That's why we're excited to share a new feature in the Catalog: Optical Character Recognition, or OCR.



PV2 Ruben Gutierrez a general mechanic, 4/21/1997, National Archives Identifier [6504747](#)

What is OCR?

OCR converts images that contain typed, handwritten, or printed text into text that can be read and searched by a computer.

Previously, records in the Catalog were only searchable based on the titles, descriptions, and other fields entered by archivists, or by tags and transcriptions entered by citizen archivists. Now, with OCR capability, text from some images in the Catalog can be extracted, making that text searchable and more likely to come up in your search results.

Currently, the Catalog's new OCR engine is applied to records in either JPG or PDF format added to the Catalog since June 2019. NARA is exploring how to retroactively process records from before that point, but right now this feature applies to millions of pages!

Here's what you can expect and how it works:

A search for "[Melvin H. Coulston](#)" returns this Bureau of Indian Affairs record with OCR data. The search term is bolded.

SEARCH TIP - If you are searching for a name or phrase, surround it in quotation marks to do an exact phrase search.

The screenshot shows the National Archives Catalog search results for the query "Melvin H. Coulston". The search bar at the top contains the query, and the results show 1 of 1 results in 0.31 seconds. The search term "Melvin H. Coulston" is highlighted in the search results. A red arrow points to the search term in the description. The description includes the title "Special Noted Files: Service Book" and the text "Melvin H. Coulston" in a red box. The description also includes the text "Card Files Relating to Indians in World War I" and "Newton, First Assistant Commissioner, Robert F. Whitehead, Chief Clerk, Philander P. Claxton, Chief Clerk, James F. Abel, 4 SERVICE BOOK". The description also includes the text "National Archives Identifier: 134530916 Container Identifier: 1 HMS Entry Number(s): PI-163 977-B Creator: Department of the Interior. Office of Indian Affairs. Education Division. Employees Section. 1911-1931".

To explore the results further, click on the blue description title. On the description page that follows, you can see the pages where your search term is found. They are listed below the description title and to the left of the image viewer:

NATIONAL ARCHIVES CATALOG

ARCHIVES.GOV LOGIN REGISTER VETERANS' SERVICE RECORDS HELP API CONTACT US About Us Privacy Policy Accessibility FOIA USA.gov

"Melvin H. Coulston" Advanced Search Login Register

Return to Search Results Results: 1 of 1 Share Export

Record Hierarchy

Record Group 75:
Records of the Bureau of Indian Affairs, 1793 - 1999

Series:
Card Files Relating to Indians in World War I, 1920 - 1920

File Unit:
Special Noted Files:
Service Book

Tag

Add a tag...

Enter new tags... Add

Comment

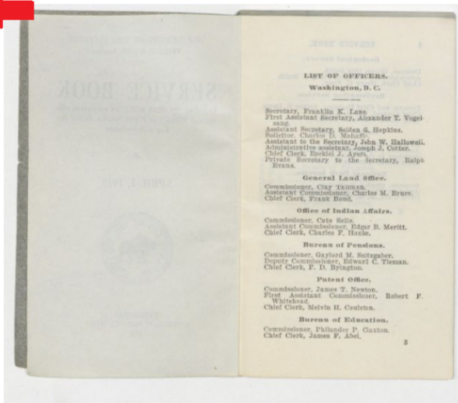
Add a comment...

Login | Policy
Need Help?

Special Noted Files: Service Book

Search term(s) found on:
Image #3

Close



View/Add Contributions

Image 3 of 74

Likewise, you may also see the page thumbnails highlighted beneath the image viewer that contain your search term. Clicking on any of the pages in the list or a highlighted page thumbnail will take you to that page.



Try it out!

You can test the capability yourself by running one of the following searches and clicking the first result returned for each:

- [“Dornier Aviation Plants”](#)
- [“Did you bail out? No”](#)
- [“Sea Devil Torpedo Attack”](#)

We still have work to do! Right now, we are investigating options to re-process items for OCR that were in the Catalog prior to June 2019. Additionally, records that are only available in PDF format currently do not provide the page jumping or highlighting capability.

OCR is not perfect! While this technology helps to make records more searchable, we still find human-entered transcriptions to be more accurate than OCR, so we still need your help as citizen archivists to transcribe records in the Catalog, and help decipher that tricky handwriting!

In case you were interested...

Technical specs: NARA's new OCR engine is powered by the open source [Tesseract](#) software. As records are added to NARA's [Amazon Web Services \(AWS\) S3 cloud storage](#), it is run through image processing powered by a series of [AWS Lambda](#) functions.

News from the Innovation Hub: 500,000 pages scanned by Citizen Archivists!

Citizen Scanners in the National Archives Innovation Hub have officially scanned 500,000 pages of historical records!

Following scanning, each digitized page is then added to the National Archives Catalog for anyone to view and download. Currently, nearly 490,000 pages of records scanned in the Hub are available in the National Archives Catalog. You can [view all records scanned by citizen contributors](#) in the Catalog so far.

Here is the 500,000th page scanned in the Innovation Hub:

3-447.

East Division
Proving No. 134,511
Department of the Interior,
BUREAU OF PENSIONS,
 Washington, D. C. *April 16, 1907*

SEE: To aid this Bureau in preventing any one falsely personating you, or otherwise committing fraud in your name, or on account of your service, you are required to answer fully the questions enumerated below.

You will please return this circular under cover of the inclosed envelope which requires no postage. Very respectfully,

Antonio Dardell
29 1/2 George Street
New Haven Conn

V. Warner
 Commissioner

1. When were you born? Answer: *Taken from China in childhood*
2. Where were you born? Answer: *China*
3. When did you enlist? Answer: *22d October 1862*
4. Where did you enlist? Answer: *City of New Haven Conn*
5. Where had you lived before you enlisted? Answer: *China born*
6. What was your post-office address at enlistment? Answer: *Clinton Conn*
7. What was your occupation at enlistment? Answer: *China Apprentice*
8. When were you discharged? Answer: *25th July 1865*
9. Where were you discharged? Answer: *City of New Haven*
10. Where have you lived since discharge? Give dates, as nearly as possible, of any changes of residence.
Clinton + New Haven to New Haven Conn 1862
11. What is your present occupation? Answer: *Tinsmith*
12. What is your height? Answer: *5 feet 9 1/2 inches* Your weight? *125*
 The color of your eyes? *Black* The color of your hair? *Black* Your complexion? *Dark*
 Are there any permanent marks or scars on your person? If so, describe them.
13. What is your full name? Please write it on the line below, in ink, in the manner in which you are accustomed to sign it, in the presence of two witnesses who can write.

Antonio Dardell

WITNESSES: { 1. *Shannon S. Thompson* Date: *April 20th* 1907
 2. *Rud S. Thompson*
 (Witness who can write his name)

PENSION OFFICE
 APR 27 1907
 U. S.

This is a form that Antonio Dardell filled out to apply for an increase in his Civil War veteran's pension. Antonio Dardell is an interesting soldier, being one of the small number of Chinese American soldiers who fought in the American Civil War.

Dardell had likely been adopted as a young boy by an American whaling ship captain, and brought to Connecticut. He enlisted in the 27th Connecticut Infantry Regiment at age 19. After the war he worked as a tinsmith, and in 1880 he became a naturalized citizen, showing his honorable discharge certificate in lieu of a Declaration of Intent. More information can be found in this ["The Blue, the Gray, and the Chinese" blog post](#).

This pension file was scanned as part of a two-week project by two citizen scanners who were interested in digitizing records relating to Asian American history. In addition to Antonio Dardell's pension, the citizen scanners also worked on his Compiled Military Service Record, and the CMSRs and pensions of other Chinese American soldiers.

Many thanks to everyone who's worked in or adjacent to the Hub now or in the past! Here's to the next 500,000 pages.

Now that these pages are scanned and added to our Catalog, help transcribe them to make them more searchable! Our latest [citizen archivist mission](#) features records scanned by citizen contributors in our Innovation Hub.

Get started transcribing!

Interested in Citizen Scanning? Learn more about the [National Archives Innovation Hub](#) in Washington, DC.

Questions or comments? Email us at catalog@nara.gov.



[Privacy policy](#)

[Subscribe](#) or [Unsubscribe](#)

Powered by [Mailchimp](#)