Using Chronicling America Podcast Series Transcript

NDNP Podcast 11: Understanding Keyword Searching

Background

In collaboration with the National Endowment for the Humanities and the Library of Congress, the National Digital Newspaper Program in Ohio developed and produced a series of video podcasts that will help teach you how to access the rich content available on the historical newspaper database Chronicling America (http://chroniclingamerica.loc.gov).

About NDNP Podcast 11: Understanding Keyword Searching

Description	Sometimes the search terms you type into Chronicling America don't always match your search results. View this podcast to find out why and learn more about keyword searching.
Hosts	Kaylie Vermillion
Duration	1:53
URL	http://www.ohiohistoryhost.org/ohiomemory/resources/tutorials

0:03	In this podcast, I'm going to explain a little bit more about keyword searching.
0:07	Chronicling America uses Optical Character Recognition, or OCR, to make newspaper pages searchable.
0:13	Basically, OCR is a computer program that reads the page and does its best to figure out what words are on it. It then creates a text document that is attached to each image and used by the search engine to find your search terms.
0:26	Let's click on an image and open the text document. The text document is composed of words and letters the OCR recognized. As you can see, there are some strings of letters that are not recognizable words and sometimes the words the OCR recognizes are not accurate and are not the words that appear on the image.
0:43	For more information about this, see the podcast titled "Overcoming Historical Language Barriers."
0:48	Because the text document underlying each image is not created by humans, errors and imperfect OCR are unavoidable.
0:55	Sometimes it has strange fonts, imperfections in the page such as creases, low contrast or minimal column dividers that throw off sentence structure which leads to unexpected search results.

1:06	If you search for more obscure news items such as a birth announcement that would only be found in one paper and the OCR missed it in that one paper, then the article will be missed altogether.
1:16	To contrast, if you searched for a news item such as McKinley's assassination and the OCR misses it in one paper, then it is likely you will get many results from other papers because McKinley's assassination was news of national importance.
1:30	Nevertheless, keyword searchability, even if it doesn't work one hundred percent of the time, does greatly increase access and is definitely time-saving when compared to working with print or microfilmed newspapers.
1:42	And that concludes our podcast about understanding keyword searchability.
1:45	Check out our other podcasts for more information about Chronicling America.