

Advantage Preservation Newspaper Archive

What does it mean when I see hieroglyphics?

When you try to open a newspaper PDF and you haven't input a search term, you're simply browsing through the newspapers page by page. In that instance what happens is the software is written in such a way that it wants to display something in the abstract. Being that it doesn't have a search to display it's grabbing whatever it can on the page and it displays it in the abstract. If you perform an actual search you'll see fewer characters.

If you perform a search and you're still seeing hieroglyphics in the abstract result, the OCR software will grab everything on the page. OCR software is set up to scan across a newspaper page. Newspaper pages are written in columns. OCR software can't distinguish this and as it scans across the page, it jumps from one article to the next across the columns.

Another cause for the hieroglyphics occurs when you search older newspapers from the 1800's and early 1900's. Frequently, these newspapers contained odd characters such as: diamonds, ovals, stars, and dashed lines to separate articles and to divide the columns and sections of the newspaper. OCR software scans those odd characters on the page and tries to display them as a search result. The same happens when the original microfilm has scratches or dust on them.

The OCR software used by Advantage Preservation is called ABBYY Fine Reader. According to reviews, It is one of the best OCR software products on the market. Most subscriber based sites like newspapers.com, Proquest, Newsbank, etc., use.

How do I get the information that is displayed in hieroglyphics?

Search using actual search terms. This will reduce the incidence of seeing those characters. Once you get to the page you want, you will be able to see the actual newspaper page with no characters. At the top of this screen, you will be given choices for downloading, printing, increasing or decreasing the page, etc.