Al-Ahram
DIGITAL ARCHIVE USER GUIDE

August 2020

WELCOME TO THE AL-AHRAM DIGITAL ARCHIVE

Founded in 1875, *Al-Ahram* (الأهرام, "The Pyramids") is one of the longest-running newspapers in the Middle East. It has long been regarded as Egypt's most authoritative and influential newspaper, and one of the most important newspapers in the Arab world, with a circulation of over 1 million.

The *Al-Ahram* Digital Archive is a remarkable collection bringing together nearly 150 years of this critical resource into one place. Previously, you would have needed to visit any number of physical libraries around the world to view these materials in a piecemeal fashion. In each library, only a small portion of this collection would have been held, and in varying degrees of condition – from musty old boxes of newsprint to long-shelved microfiche. This new digital collection assembles the massive corpus of issues of this newspaper into a single web presentation for your ease of access.

The *Al-Ahram* Digital Archive contains all obtainable published issues of the newspaper from 1875 to present, with an additional year's worth of content added on an annual basis. This archive offers scholars the most comprehensive collection available for this title, and features full page-level digitization, complete original graphics, and searchable text.

TOPICS COVERED IN THE USER GUIDE

- **Searching the Archive**
- **Access and Account Set-Up**
- **Interface Display, Navigation and Tools**
- **OCR Text Correction**
- **Additional Resources**

HAVE MORE QUESTIONS?

Please contact us at online@eastview.com.

east view
INFORMATION SERVICES

Uncommon Information
Extraordinary Places

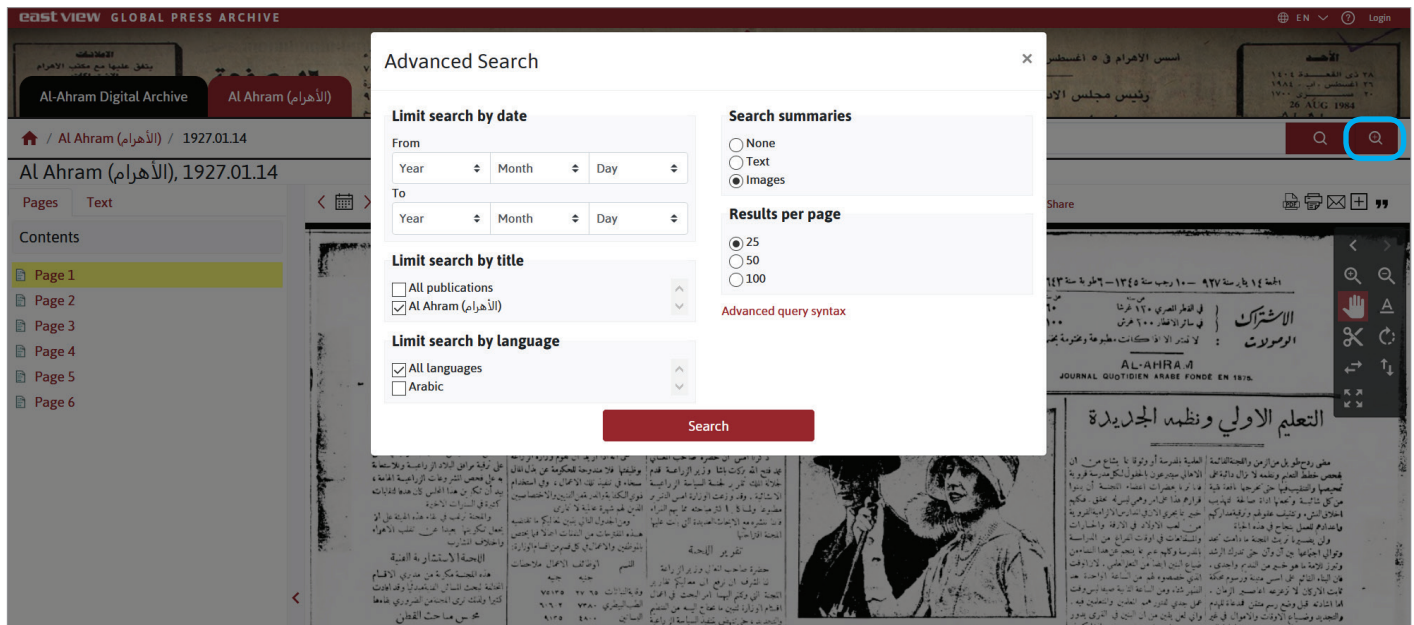## SEARCHING THE AL-AHRAM DIGITAL ARCHIVE

Basic Search

The basic search bar is located on the header at the top of every page. Input your search term(s) and press "Enter" or click on the magnifying glass icon. Search terms may be entered in Arabic or English; users cannot search using transliteration.
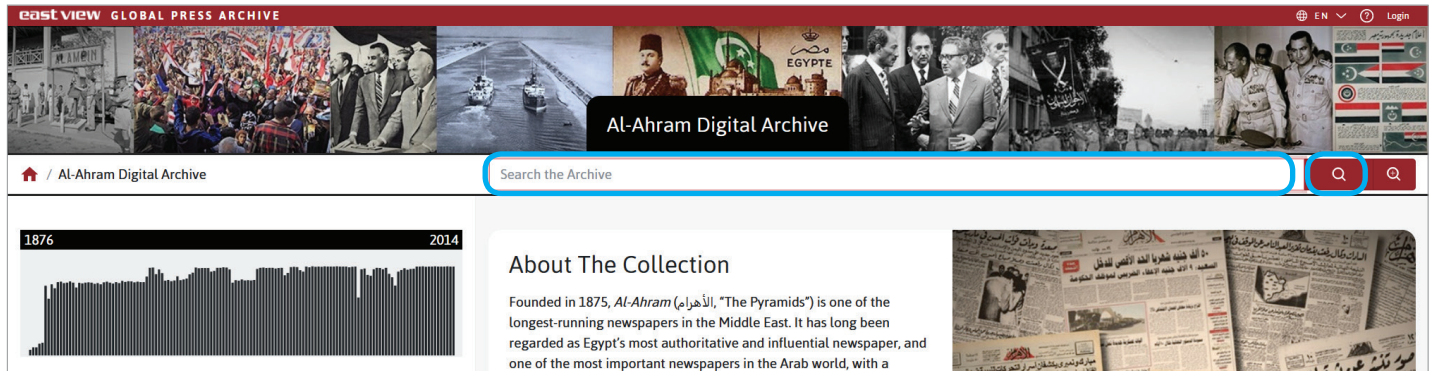
Advanced Search

To perform an advanced search, input your search term(s) into the basic search bar and click on the magnifying glass icon with a "+" sign. A pop-up window with advanced search options will appear on the screen (see below). Options are available to limit a search by date, and to select how search results are displayed and number of results per page. A link to the Help section on advanced search query syntax is provided for reference.

Al-Ahram Digital Archive User Guide

Search Results

Search results can be displayed in three views (see examples above; select your preferred view from the Advanced Search menu):

1. none – displays page image, metadata and URL
2. text – displays search query highlighted within text
3. images – displays search query highlighted within image of text

Search results can be further refined by date range using the Date filter on the left panel. Results can also be sorted by relevance or date (oldest first or newest first). Use the red star icon to add results directly to a private list stored in My Account (see **pages 4-5** of this user guide for more on private lists and My Account).

Please note that clicking on a search result will take you to a view of the newspaper page containing that result, and not to a specific article on that page.

Search Tips

Search

Searching without quotation marks will feature results based on simple word count occurrence. As an example, searching for الحزب الوطني (without quotation marks) will tend to locate pages containing either word anywhere on the page.

Exact Search

As a rule, using quotation marks (") around the search query will result in more targeted search results. Using quotation marks limits the search result to words appearing in that exact order in the text. The best-targeted approach to search is to use quotation marks as this provides results matching exact phrases in the digital text. See the example of search results, above, which shows results for the search query "حدود مصر" (with quotation marks). Results are limited to pages that include this exact phrase.

Wildcard Search

An asterisk (*) can be added to a search term to add a wildcard factor of multiple letters following the letters in succession of the intended target. For example, searching for حدو* will locate occurrences in the digital text which include items such as حدودها, حدوث, and حدوثه and of course, حدود.

Proximity Search

A proximity limiter can also be added using the tilde (~) at the end of a phrase. For example, واشنطن رونالد ريجان ~7 will locate page instances in which the proximity of any two of the words is equal to or less than a distance of 7 words away from any other word in the query.

ACCESS AND ACCOUNT SET-UP

Web Address: https://gpa.eastview.com/alahram

Access to the *Al-Ahram* Digital Archive is available via the authorized IP network of a subscribed institution. If you are unable to access content within the *Al-Ahram* Digital Archive, please contact your institution for assistance with IP access.
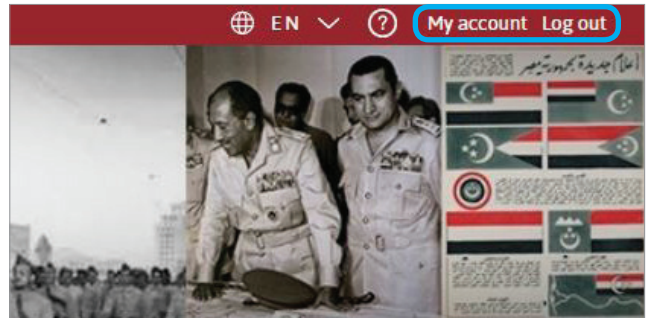
My Account

Individuals accessing the *Al-Ahram* Digital Archive through an authorized IP network of a subscribed institution can opt to register a personal account. Establishing a personal account to the archive is a convenient feature that allows researchers to download content, make OCR text corrections, and build a library of saved content, searches, and more.

Registration and Login

To set up a personal account, you will need to access the archive from your subscribed institution IP network, then register your personal profile using the Login link in the top right corner of the interface. Follow instructions to complete the registration form. Once registered, login with the username and password you provided to access your account during future sessions in the archive.

When logged in, you will see a "My Account" link in the top right corner. Use this link to change your password, change your registration email address, access your saved private lists and notes, and opt in or opt out of storage of your history of searches and views, etc.


> **NOTE:** Your personal account is tied to the account of your institution. Access to your account is only available when you are accessing the archive through the authorized IP network of your institution.

Creating Private Lists and Notes in My Account

The *Al-Ahram* Digital Archive allows for registered users to create private lists of page images for future reference. To save a page image to a private list, click on the "+" icon in the upper view area of any page view.
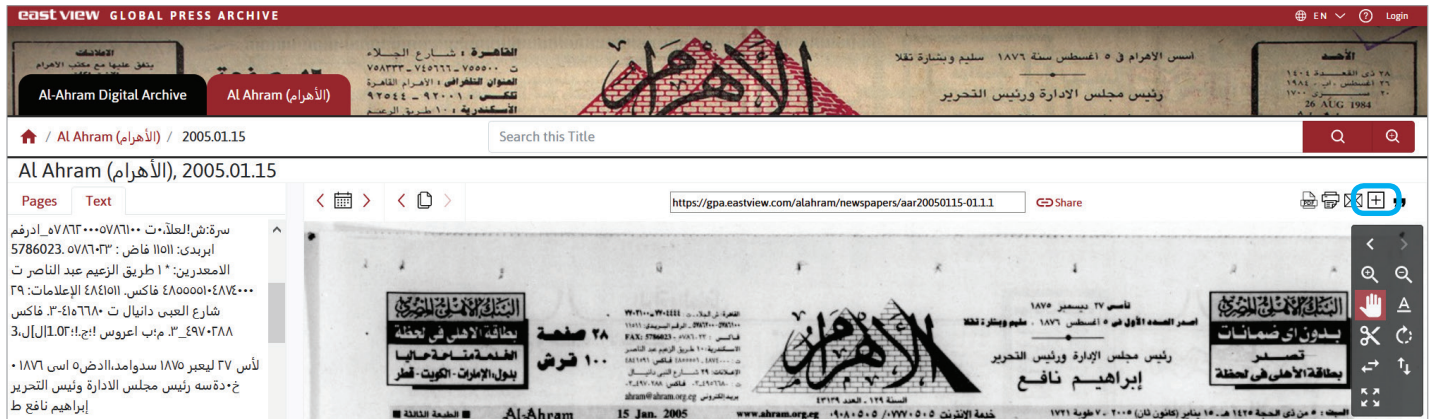
Private lists can be annotated with notes or reminders pertinent to the pages you select, and you can create multiple lists. Metadata is automatically added to your selected content so that you don't need to keep track of dates and page information.

Private lists that you create are available in the My Account section of the website. From this area, you can view and manage your lists, as well as email your list to yourself. Note that all lists and annotations created in My Account are kept private to the user and are not viewable by other users.

Al-Ahram Digital Archive User Guide

INTERFACE DISPLAY, NAVIGATION AND TOOLS

Home Page View

The *Al-Ahram* Digital Archive home page is comprised of a collection description, search tools, and navigation options for viewing content. Select features are described below.

Interface Language

Select the language dropdown from the top menu to toggle between English and Arabic language interfaces. This may be done at any time from any page in the interface. Functions of the interface remain the same regardless of language selection. See the **Additional Resources** section of this user guide for information on changing your device's input language.

Login/Register

Login or register for My Account using the link here. Once logged in, this area will display links to access your account and to log out.

Search Tools

Use the search field to conduct a basic search of the *Al-Ahram* Digital Archive, or select the magnifying glass icon with the "+" sign to go to advanced search options. Detailed search instructions and tips are included on **pages 1-3** of this user guide.

Help

Select the "?" icon in the top menu bar to visit the Help page, which includes information regarding searching, advanced search, advanced query syntax, OCR (Optical Character Recognition), how to correct OCR text, printing, technical requirements, and use of the materials, citation and copyright.

Al-Ahram Digital Archive User Guide

Home Page View, continued

Title Navigation

Clicking on "Title Navigation" directs you to the main landing page for the newspaper. This landing page displays full metadata and, like the main archive home page, also includes Date Navigation and "On This Date in History" options for navigating to content within the archive (see further details on these features, below).

Date Navigation

Use the Date Navigation feature to quickly select the year, month and day of the issue you wish to view. To use, select the year and month from the dropdown selection boxes. Months that include content are indicated by a "(+)" next to the month name. Once the year and month are selected, the calendar will display days for which content is available with a dark grey box. Click on the day to view the issue.

On This Date in History

Each time you visit the *Al-Ahram* Digital Archive, this area will fill with randomly generated selections from different years of the archive corresponding to the current date. A thumbnail image displays a small part of that issue's front page, with the date of the issue given to the side. Please note that this thumbnail image is also randomly generated to capture different parts of the newspaper page, so you may see a portion of a headline, a story, a photograph or even an advertisement. Click on any of the random selections shown to view the full newspaper issue.

Issue Page View

Toggle between "Pages" and "Text" views in the left panel. "Pages" view lists the individual pages of the issue, which are hyperlinked to jump directly to the selected page. "Text" view (seen above) displays OCR text of the page being displayed.

Use the arrows on the calendar icon (left) to navigate to the next or previous issue. Clicking on the icon itself will lead to the main landing page for the newspaper, where date navigation can be used to browse the publication. Use the arrows on the paper stack icon (right) to navigate to the next or previous page within the issue currently on view. Clicking on the icon itself will automatically display the "Pages" view in the left panel.


This bar displays the unique issue-level permanent URL for each issue in the archive. Use the "Share" icon to copy the URL to your local clipboard.

 From left to right:

- Save the page currently displayed to your local device as a PDF
- Print the current page via your local device
- Send an email containing the date of the issue currently displayed
- Save the current page to a private list stored in "My Account" (see **pages 4-5** of this user guide for more on private lists and My Account)
- Generate citations in MLA, APA, Chicago or RIS format
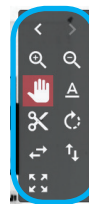
 From top row to bottom row:

- Navigate to the next / previous page within the issue currently on display
- Increase / decrease magnification of the page currently on display
- Grab and move the page currently on display / select text from page to copy and paste
- Clip selections from page to view, copy or save / rotate image (rotates 90 degrees with each click)
- Fit image display to width of page / fit image display to height of page
- View image in full screen display

OPTICAL CHARACTER RECOGNITION (OCR) TEXT CORRECTION

Content in the *Al-Ahram* Digital Archive stretches over nearly 150 years and has been obtained from a wide variety of sources, making the uniform presentation of newspaper issues all but impossible. In producing the archive, East View has taken great care to locate the best possible representation of each issue of the newspaper including, in some instances, the use of other editions of the publication. However, it is important to recognize that certain limitations mean that there is a range in the quality of images throughout the archive, which in turn affects the accuracy of OCR and the resulting searchable text. Please refer to the box below for examples of factors that impact OCR.

All these factors present challenges for text recognition. For instance, low-quality paper, poor imaging conditions, and aged microfilm in some cases combine to render portions of an image completely dark. Even in less extreme instances, the factors below can complicate OCR, especially when dealing with a language like Arabic, where letters with similar characteristics such as dots placement (such as ق and ت , particularly in medial form in the middle of a word) can be confused for each other. Likewise, in the case of small typefaces, there can be confusion in the reading of letters such as ع and م in initial character position in a word. Depending on how tiny the typeface is on the page, OCR could recognize both these characters as the latter Arabic letter due to its "closed circle" appearance.

The *Al-Ahram* Digital Archive platform allows for registered users to help improve the accuracy of OCR within the archive by correcting instances in which the OCR text is incorrect. Read on below for instructions on how to perform OCR text correction in the *Al-Ahram* Digital Archive.

FACTORS AFFECTING IMAGE QUALITY AND OCR

IMAGING OF NEWSPAPER PAGES – Lighting, focus and camera quality of the time period in which the original newspaper pages were photographed can have a tremendous impact on the quality of the image available in the archive. In general, archival actions taken with older material and photographed for microfilm using technology before the 1980s leaves much to be desired from a modern perspective. Likewise, the eventual aging of the microfilm itself can also result in dark or blotchy images.

PAPER QUALITY – The quality of the paper available to the Egyptian press was often itself poor, and this resulted in pressed ink often immediately smearing at the time of printing, a very evident problem with many Arabic newspaper publishers even into the 1990s. In many cases, attempts were made to use Western paper for the front page of an issue, whereas the remainder of the issue was printed on Soviet-supplied paper – paper which was far more prone to smearing and aging poorly.

FONT SIZE – The use of particularly small typeface can render some text unrecognizable to OCR. In Egypt, efforts to conserve paper led to the use of truly tiny text on the page, resulting in many lines of text being nearly unrecognizable. In such instances, it is not uncommon for dots to be smeared or morphed into dashes, or for disconnected letters in Arabic to become "connected," thus lowering the quality of OCR.

HANDWRITTEN TEXT – During some eras of the newspaper, *Al-Ahram* was produced by hand scribes solely. In other eras, the newspaper employed a mixture of handwritten script – particularly for headlines – as well as typeset Arabic. In general, errors in OCR are more likely when deciphering handwritten script than typeset characters.

OCR Text Correction Instructions

The OCR text for any newspaper page in the archive can be displayed by selecting the "Text" tab on the left panel. To correct the OCR text, first make sure that you are logged in to your personal account (see **pages 4-5** of this user guide for more information on personal accounts). Next, click on "Correct this text" on the left panel.

When you click on "Correct this text," you will be presented with a view of each page in the issue. You can then select any zoned block to edit the OCR text. A zoned block is a rectangular selection where the machine reading has determined that there is a line or block of text. In the image below, the light section of text in the middle is the zoned block of text being edited, with the corresponding OCR text presented for editing on the left panel.

To edit text, place your cursor in the text box on the left panel that contains the text you wish to correct. Text can be edited line by line. After correcting a piece of text, click "Save."

You can continue to make corrections to the same line of text, to other lines of text in the same block, or move on to the next block by clicking "Save & next," clicking the "Next" button, or selecting another block of text from the image in the right panel.

Once all corrections are done for a given block of text, click the "This Block is completely correct" box, located at the bottom of the text boxes. Your changes will take effect immediately and will be available to all users of the platform.

OCR Text Correction Example

In the example to the right, we can see that the OCR has misread the Arabic number for "200" as "20*." In addition, the OCR has picked up on some degradation of the original ink and paper and has placed an errant semicolon at the end of the phrase. Placing the cursor in the text box, fix the Arabic numeral and delete the errant semicolon. With all corrections made, click "Save" and then "This block is completely correct."
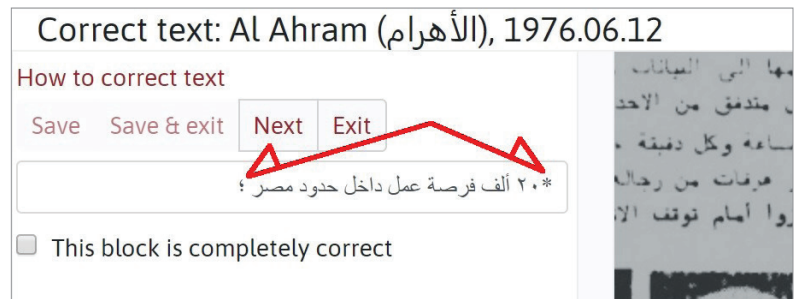
Now that the text has been corrected, the section that was edited appears with a yellow background, indicating to future users that this text has already been corrected.
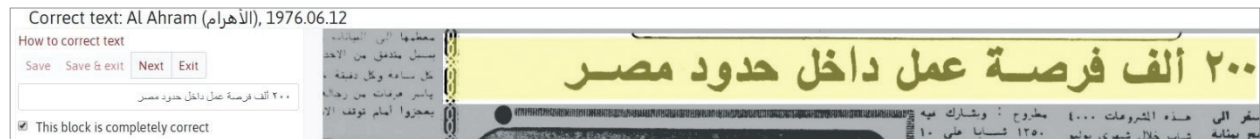
To exit the text correction view, click the "Return to viewing mode" link located in the top right corner of the page. Likewise, if you click "Save & exit" instead of "Save," your changes will be saved and you will be returned to the normal page view of the newspaper.

> **NOTE:** Additional instructions and tips for OCR text correction are available from the platform host at https://veridiansoftware.com/knowledge-base/user-text-correction-guidelines/

Personal Correction Statistics

If you have made any corrections to the OCR text, the system will keep track of your edits, as well as allow you to see a display (by user Display Name) of other individuals who have also contributed towards OCR text correction. In the My Account section, click on the "Lines of text corrected ... More information" area. This will display statistics about your editing and even includes a "Text Correctors Hall of Fame." Please note that you can edit your own "Display Name" to control what other text correctors see in this display.

ADDITIONAL RESOURCES

Arabic Keyboard and Language Tools

The *Al-Ahram* Digital Archive interface supports input from your keyboard when it is in a setting of Arabic characterization. For Windows, simply change the keyboard input language from ENG to Arabic (generally represented by symbol ع). You can switch back and forth between ENG and ع as needed. The same can be done on the Mac operating system using the Apple menu > System Preferences > choose Keyboard layout. For further instruction, see:

For Microsoft Windows
https://support.microsoft.com/en-us/help/4496404/windows-10-manage-the-input-and-display-language

For Mac OS
https://support.apple.com/guide/mac-help/type-language-mac-input-sources-mchlp1406/mac

Online Arabic keyboard
https://www.lexilogos.com/keyboard/arabic.htm

Other Arabic-Language Resources

Middle Eastern and North African Newspapers East View Information Services; an Open Access collection
https://gpa.eastview.com/crl/mena/ and https://www.eastview.com/resources/gpa/crl-mena/

Arabic and Middle Eastern Electronic Library (Near East Collection) Yale University Library
https://web.library.yale.edu/digital-collections/arabic-and-middle-eastern-electronic-library

ICON Database Center for Research Libraries; an index of newspaper and microfilm collections of newspapers worldwide
https://www.crl.edu/archiving-preservation/print-preservation/icon-database

Arabic Collections Online New York University
https://www.crl.edu/archiving-preservation/print-preservation/icon-database

MENAdoc Open Access Collections University and State Library Saxony-Anhalt , Germany
http://menadoc.bibliothek.uni-halle.de/

The *Al-Ahram* Digital Archive is a part of the East View Global Press Archive®, which is the result of a landmark initiative of Stanford Libraries and the Hoover Institution Library & Archives to digitally preserve and make more accessible thousands of original print newspaper publications collected by the Hoover Institution and now housed by Stanford Libraries. For more information, visit eastview.com/gpa.

Want to participate in Global Press Archive?

East View is interested in collaborating with academic institutions and publishers to preserve and provide access to their collections. For further information on how you can participate in Global Press Archive, contact gpa@eastview.com.