

## Handwritten Arabic Dataset Arabic-Handwriting-1.0



**Arabic Dataset  
Available for  
Research &  
Development**

**Announcing the first full page, PAW level ground truthed, Arabic collection for handwriting research! This novel collection provides a unique set of documents produced by various authors, in different formats, sizes, and styles. Contents were transcribed on a variety of papers, at different writing “speeds,” using different writing instruments.**

<p><b>Description of Image Collection</b></p> <ul style="list-style-type: none"> <li>• 200 unique documents</li> <li>• 5000 handwritten pages</li> <li>• A wide variety of document types: diagrams, memos, forms, lists (including Indic and English digits), poems</li> <li>• Documents produced by various writing utensils: pencil, thick marker, thin marker, fine point pen, ball point pen, black and colored</li> <li>• Available in binary and grayscale</li> </ul>	<p><b>Ground Truth</b></p> <ul style="list-style-type: none"> <li>• Flexible XML formatting</li> <li>• Author Metadata             <ul style="list-style-type: none"> <li>◦ Writer ID, Geographic Origin, Age, Gender, Education, Arabic Training</li> <li>◦ Page Level</li> <li>◦ Text Line Level</li> </ul> </li> <li>• Image Metadata             <ul style="list-style-type: none"> <li>◦ Word Level Bounding Boxes</li> <li>◦ Part-of-Arabic-Word (PAW) offsets</li> <li>◦ Word Content (UTF-8)</li> </ul> </li> </ul>
<p><b>Tools</b></p> <ul style="list-style-type: none"> <li>• Image Viewer to browse images and metadata</li> <li>• XML Parser to parse groundtruth files</li> </ul>	<p><b>Cost</b></p> <ul style="list-style-type: none"> <li>• Academic: \$500</li> <li>• Industry: \$1500</li> <li>• Government: Contact Us (see below)</li> </ul>



### Licensing

This data set is being distributed by Applied Media Analysis, Inc who retains all rights to the scanned images and non-content metadata in this collection. AMA provides this data on an "as-is" basis and makes nor representations or warranties regarding the correctness of the data and disclaim liability for damages resulting from its use. We cannot provide unrestricted permission regarding use of the content of this data, as some data may be covered by other rights. Only the images are being distributed and they are being distributed for research purposes only.