## Project Home in Google Home & Check out from SVN

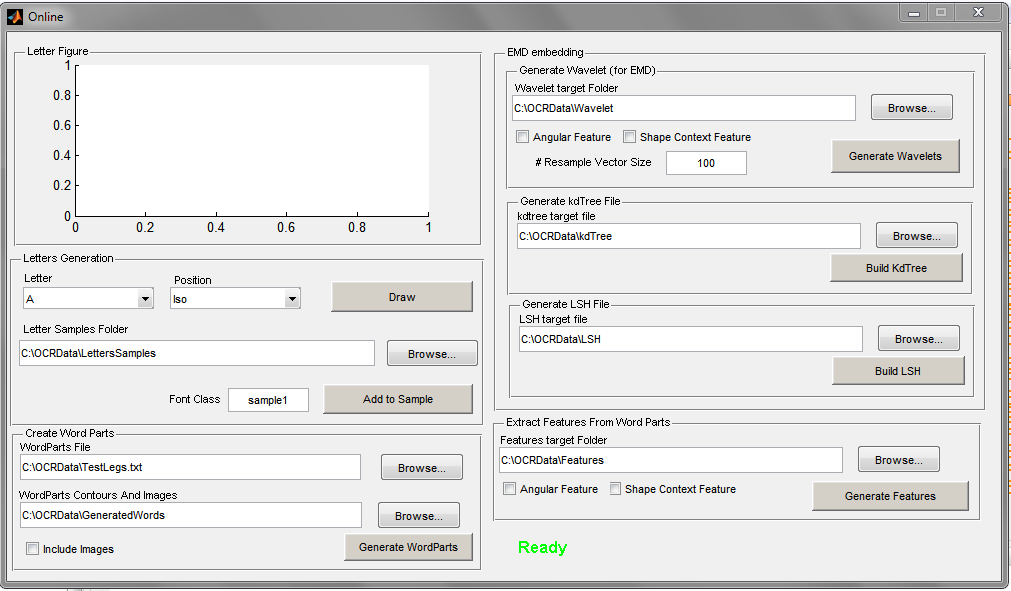
1. The Project is stored and managed in google code project, following the link to the project Home - http://code.google.com/p/arabic-icr
2. To checkout the code you need to download tortoise SVN and use the link is in the Source tab in the project Home. (https://arabic-icr.googlecode.com/svn/trunk)
3. If you find any issues/defect please create a new issues in the Issues tab.
4. After Checking out, you will get 3 directories,
   1. Code – contains the application code
   2. Data – contains the Adab database, the letters samples that I have collected and WordParts, a text file that include all the Arabic word parts.
   3. Knowledge – Contains documentation about the recognition process and some documentation.

## Handling Folders

1. Create a folder named OCRData under C:\ (C:\OCRData)
2. This folder will contain all the output data that the application generated. (Word Parts folder, Wavelets folder, Kdtree, etc…)
3. Copy the LettersSamples folder that is located under the Data folder to C:\OCRData.
4. Copy the file TestLegs.txt that is located under Data\WordParts to C:\OCRData

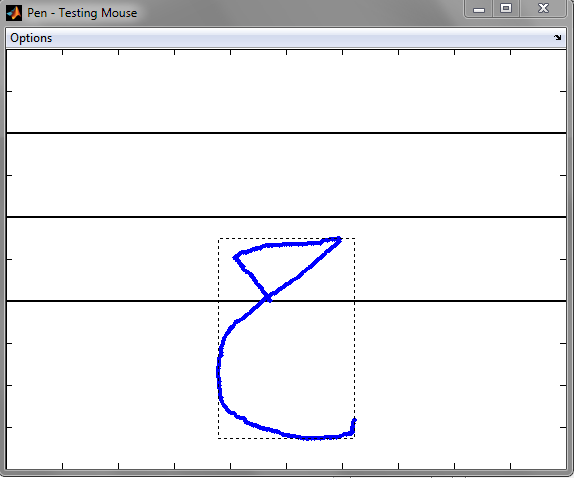
## Online Application’s UI

1. Open matlab.
2. The main UI matlab file is online.m.
3. To start the application UI run the function online.m that is located under Code directory.
4. Online.m looks as follows:



## Collecting Letters Fonts

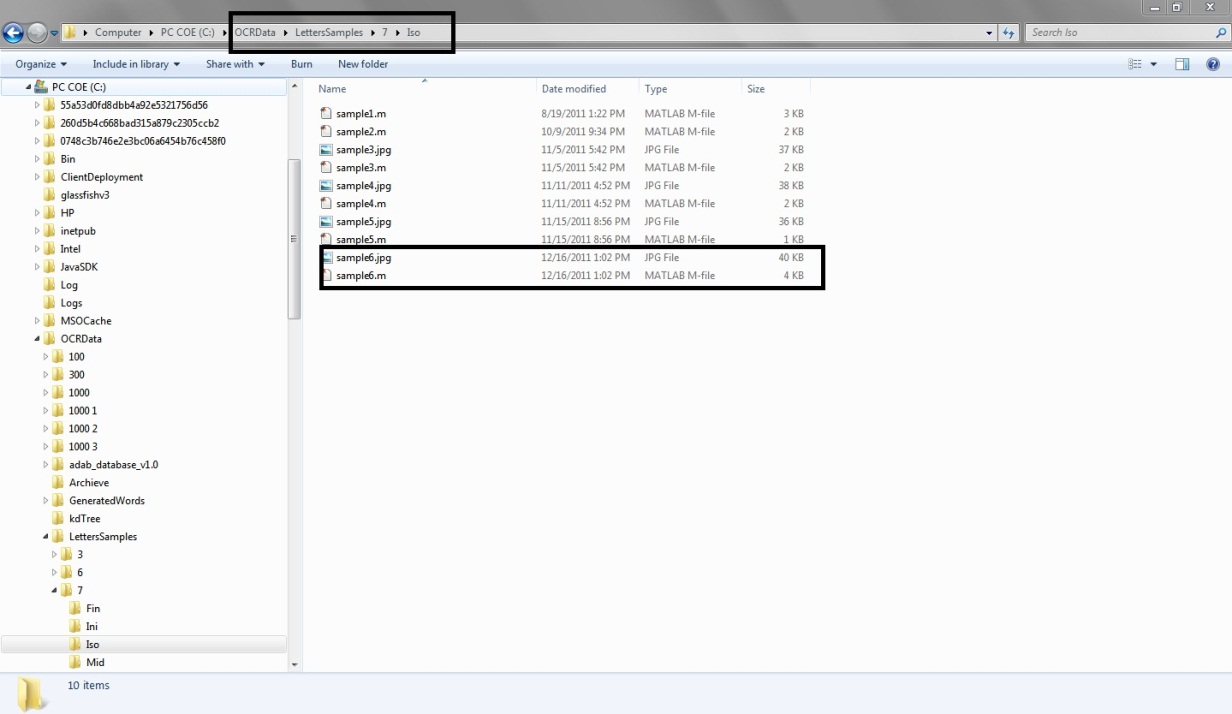
1. Select the Letter and the position and click Draw.
2. A mouse tracker windows open:



1. Draw the letter in one stroke and then close it.
2. The letter will appear in the axes in the online window.
3. Choose the FontClass and then click add to sample.
4. I have already collected 5 FontClasses (“sample1” to “sample5”) so start from FontClass named “sample6”. Please compel to this naming convention.
5. When clicking add to sample 2 files will be created under the corresponding folder (i.e. in our example under C:\OCRData\LettersSamples\7\Iso): 1. “sample6.m” and “sample6.jpg”

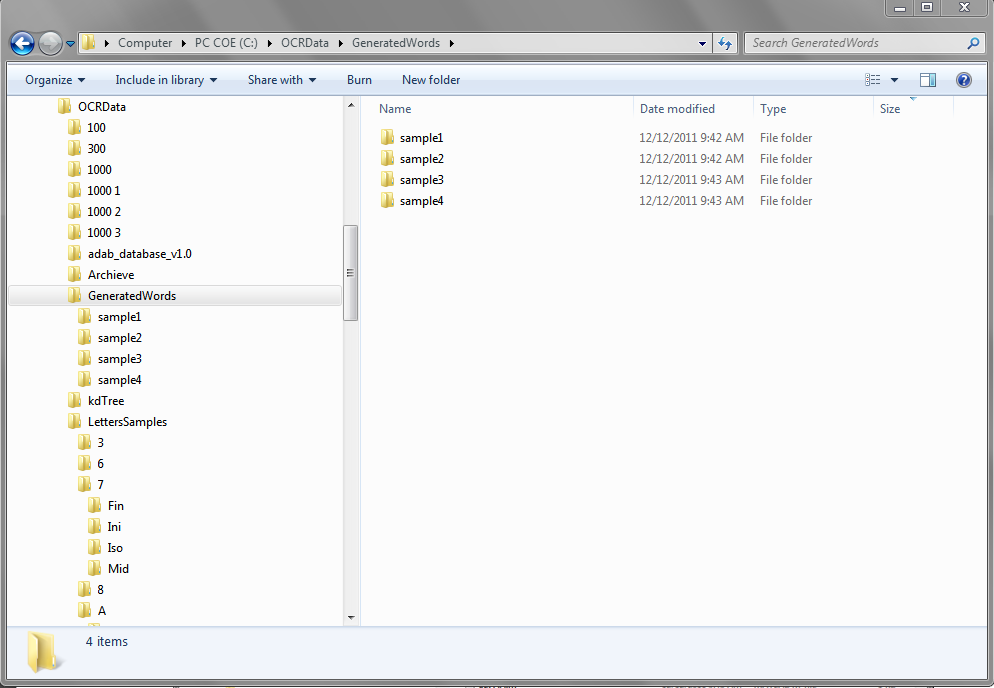
The first contains the letter sequence and the latter the letter image.

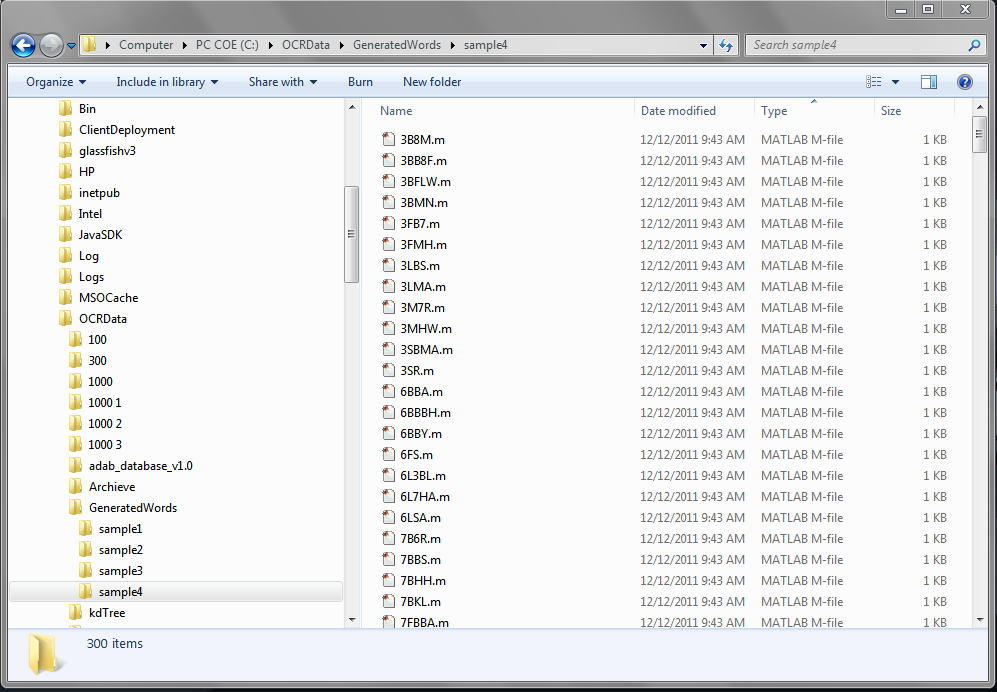
1. It is very important to specify the correct FontClasse and not to override the font classes I have already creates (sample1 to sample5).



## Generate Word Parts

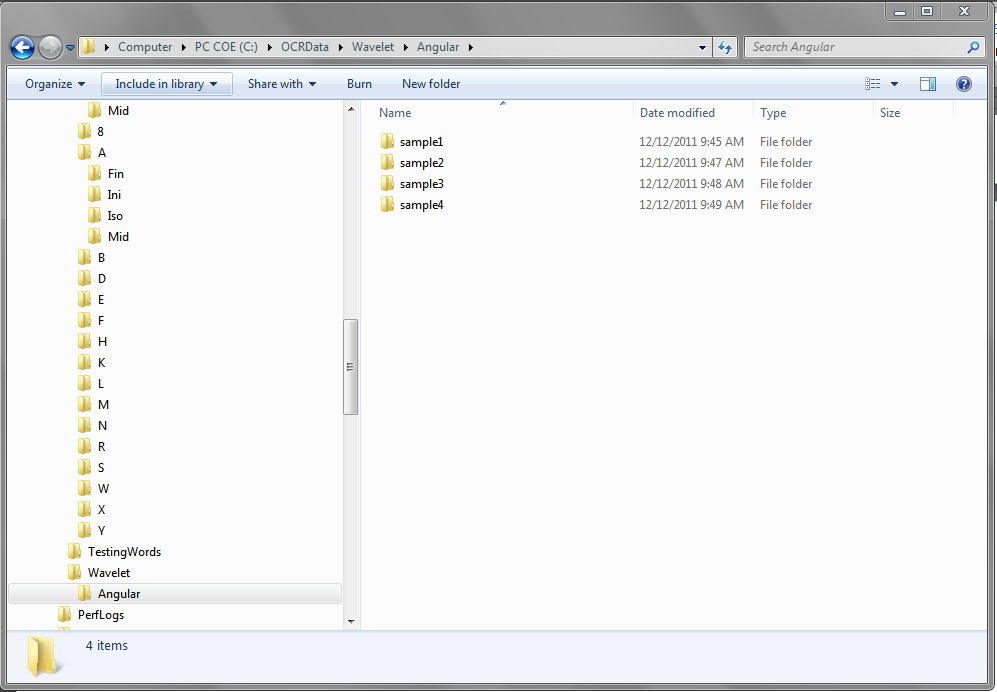
1. After creating a font class you need to create a parts sequences based on the letters you have drawn.
2. This will be done using the Create Word parts panel.
3. First you need to choose which word parts you create, thus choose the file that contains the word parts you need to create. The file file TestLegs.txt is the default but it contains the all the Arabic wordparts ~22k and to create all the sequences will take hours. so you need to create a file that contains a subset of the content of this file.
4. There is a matlab function I wrote that create a file with a given amount of wordparts and it create a file with a random subset of this file. It is located under Code/Utils and its name is “GenerateRandomizedWPFile”
5. When Clicking the “Generate WordParts” button a directory will be created in the path specified in the “WordParts Contours and Images” Textbox and it will contain the word parts that the “WordsParts File” specifies based on the Letters class specified in the “Font Class” textbox in the “Letter Generation” pane.
6. If you choose to “Include Images” The image of the word will be added to the output folder.
7. The output folder will look as follows:





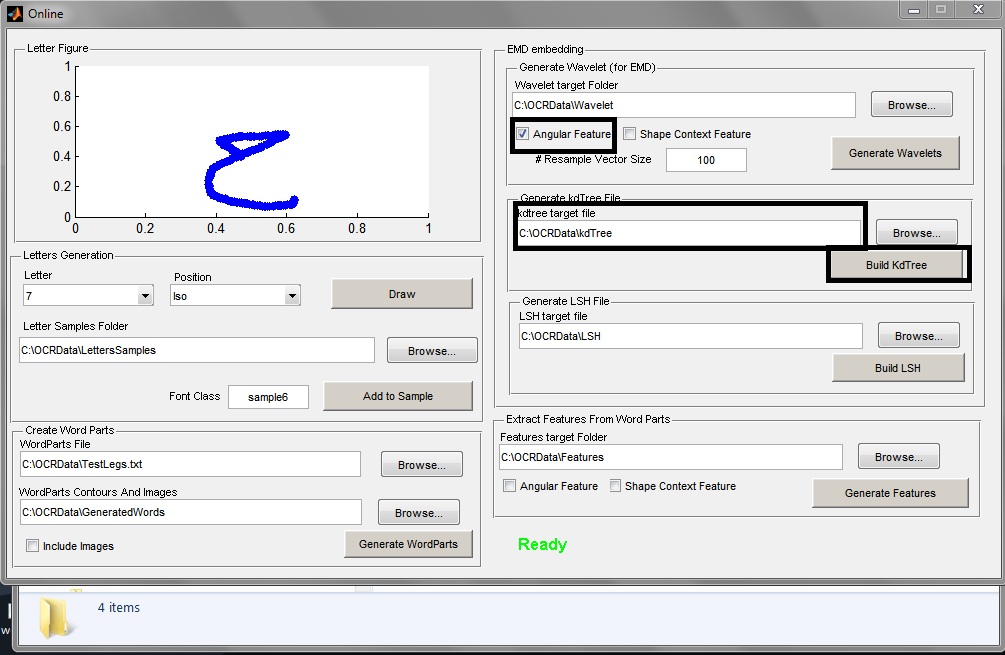
## Generate Wavelets

1. Choose the output folder in “Wavelet Target Folder” TextBox
2. Choose whether to create wavelets for the angular feature (MAD) or for Shape Context or both.
3. Click “Generate Wavelets” button.
4. The wavelets will be created from the “WordsParts Contours and images” folder.
5. Do not change the “Resample Vector Size”.
6. The wavelet folder will look as follows:



## Kdtree data structure Generation

1. After creating the wavelet embedding you will need to create the kdtree data structure.
2. Just specify the output directory and click the “Build Kdtree” button.
3. It is important to specify on which feature wavelet the kdtree will be created from, the Angular or the shape context or both.



1. The output file will be named according to the feature selection and the kdtree directory will look as follows:

