**Shecodes – Final project, Adi Rosenthal  
Hebrew handwrite to machine encoded text**

**Link for git project :** <https://github.com/Adirosenthal540/shecode_finalProject>

**Goal**

The goal of this project is implementation of a software that would be able to convert a picture (handwritten or printed text) in Hebrew to machine encoded text.

The project divided into three main processes:

1. Create processed handwrite data in Hebrew. Should meet sufficient quality data with high quantity.
2. Trained OCR natural network – tesseract.
3. Produce Graphical User Interface

**How does the software work?**

**Main.py:**

The code part that connect with the user and contains the GUI.

**Controller.py:**

Getting the information from Main and defines what process should be run.

**ModelTesseract.py:**

Use for export text using tesseract and has functions using to check trained networks.

**DataManager.py:**

The only component who has connection to database. Can insert, delete and read data from it.

**HandwrittenDoc.py:**

The component who responsible for extract data – image & txt files, from scanned documents.

**ImageProcessing.py:**

Contains image processing functions and the class “ImageProcessing” who saves image’s data during the software running.

**Config.py**

Save the home directory, the user needs to set the dictionary before the software run.

**More useful codes:**

**Check\_tesseract.py:**

The file call function Check\_model\_tesseract and find accuracy using the validation folder. Need to defined the model name.

**Make\_fonts\_data.py:**

The file contain function to automate the process of converting new font of handwriting to data for the training process

**Operating instructions:**

1. Need to clone the project from GitHub
2. Install all the python classes indicated in the file – “Requirements.txt”
3. Install tesseract 4
4. Open the config.py file that in the code folder and change the home\_directory to the folder of the project in your own computer.
5. Run the program 😊