User Manual

Program description: This program is used to recognize text from an image, handwritten or printed, in Hebrew and English. The program works only on pure text images, without pictures, and on plain paper without lines. The program uses neural networks to recognize letters.

The Menu:

Filtering:

- **Rotate**: Rotates the image to be perpendicular. For optimal rotation the image must be at an angle of no more than 40% from perpendicular.
- **Clean**: Median filter to remove noise.
- **Threshold**: Turns all text into black or white only.
- **Line straightening**: Lines that are written sloped would not be as well recognized as straight lines. This straightens line which improves recognition. Note that the lines must still be sufficiently separated for this to work properly.

Font:

- **Create Font**: Can create a font from a text image, an empty font, and a font from a list of system fonts.
  - From image: Displays the image and you can enter each character in the corrections window.
  - Empty: Creates an empty font.
  - From list: Choose the font you want to create.
- **Deploy Dataset**: Creates images from all the characters in the dataset in the ims folder.
- **Train Network**: After creating a new font or adding corrections the font network must be trained on it. You can choose the number of training cycles and the training error. It's advisable to keep the default values.
- **Create Handwriting Font**: The "Create Font" menu is only used for computer fonts. For Handwriting fonts you must use this menu.
  - Print Input Pages: Choose3 your printer and print the 6 pages. Fill out each one by writing in the black square the corresponding character. Then scan the result and save it as images.
  - Learn From Pages: Choose the 6 pages you scanned. Your handwriting will be learned by the program. You must train the network before starting to recognize images.

Scripts:

- **Create all fonts**: Create the most popular system font networks.
- **Train all fonts**: Train all fonts in the program.

Server:

- Update the program and the server with the latest versions of the networks.

The Buttons:

- **Select Text**: Select an image or PDF file containing the text you wish to recognize.
- **Get Text**: Recognize the file. You will be asked what language the text is in. If you didn't choose a font, you will also be asked whether the text is handwritten or not.
- **Spell Check**: Spell check the result.
- **Add Correction**. Choose a wrongly recognized character, enter the correct character and press this button. You will need to train your network after entering the corrections.

Right click menu in text area: If you mark a word and right click it you can add the word to your spellchecking dictionary.

How to fill out learning pages:

- Use black pen. If you use blue or any other color pen, increase the contrast in the scanned image. But black pen is recommended.
- Keep the pages clean. Any spots, or dirt will interfere with the recognition.
- Keep within the square borders.
- Bigger letters should be bigger in the square. Smaller letters should be smaller. If a letter is bigger than another in the text, it should be bigger in the square.

How to write text:

- Use black pen. If you use blue or any other color pen, increase the contrast in the scanned image. But black pen is recommended.
- Keep the pages clean. Any spots or dirt will interfere with the recognition.

How to add corrections:

- Choose on the text the letter you wish to correct, see its image I the small square on the left. Write the correct character I the widow left of the letter.
- If you see in the letter image that it is different from expected, for example small and/or located on the wrong side of the line, it's not recommended to correct it as it would interfere with future correct recognition.
- If the letter "א" is recognized as 2 letters it should be corrected as "וט". If the programs finds "וט" in the text it will ask the user if it's "א". (In case the user answered yes the recognition of "א" will improve but there will be problems with correctly displaying the letter in the repair window).

Improving spell checking:

- The spell checking mechanism include the hunspell dll library which supplies the suggestions. It uses the files *.dic and *.aff in the Nhunspell folder. It is possible to modify them to get better suggestions. The manual for that is included in the documentation folder.
- The best suggestions are chosen using a weights matrix contained in the "visual similarity.txt" file in the Nhunspell folder. This text file was created from the "visual similarity.xls" excel file. It is possible to open and edit the txt file in excel or open the xls file, edit it and save it as a Unicode text file. The file name must always be "visual similarity.txt" and it must be in the folder Nhunspell.

Server:

- **Settings**: will open the settings window with two options "Update server host" and "Port" (Default settings is *hocr-server.no-ip.org*, port: 5656).
  After press on "Save settings" button, program will check entered values and save it in serverset.txt file
- **Show client networks:** Will show current font networks which exist in clients`s folder, with all information about this networks.
- **Open updates log file:** Will open log file in default program in Windows for txt files.
- **Update Fonts:** Will open update window, where you can press "Start update" button for star update. The current status and process will appear in log window and will be saved in log file. "Clear log" Will clear all text in the log window (but will not clear it in the server log file). After update to close this window press "X" button.

Server manual:

Program description: This program is used to make available to do updates of fonts networks files of HOCR program. This is separate program from HOCR. It has own networks files which can be stored in chosen folder. For make it comfortable to do update for client you can use registered host on www.no-ip.com.

You can manually enter this site an enter user information :

Username: hocr.project@gmail.com

Password: haifa2012

The host name is: *hocr-server.no-ip.org*

Or you can download client which will automatically update your host to current IP address.

If you have router or more complex network hardware you must to do port forwarding to your machine.

To start Server run ServerHOCR.exe file from Server\Server\bin\Debug folder.

The tools Menu:

Show server`s networks:

- **Will show current font networks which exist in server`s folder, with all information about this networks**

Open log file:

- **Will open log file in default program in Windows for txt files.**

Settings:

- **Will open window where will be shown current settings of the server.**
- **Server port**: current port on which server will be run or running.
- **Server folder**: current folder of server.
- **Max file size (in bytes)**: Maximal size of the file the server can work with (current size is 10 mega).
- **You can press on Load defaults**: it will load values from program code.
- **When you press "save settings"** the program will check the entered information and will save it in file

Start server button:

- **Will start server and will wait for clients connections**
- **All current action will appear in log window and will be saved in log file.**

Stop server button:

- **Will stop the server.**

Clear log:

- **Will clear all text in the log window (but will not clear it in the server log file).**