Text4Less

Nick Redmond

NEUMONT UNIVERSITY

2013

Abstract

This document serves as the formal proposal for my winter 2014 Capstone project. The project is an application in optical character recognition, which is the automated process of reading text from images and converting that text into a computer-readable format. As a proposal, this document explains the purpose of the project, and it outlines the functions that the final product will be capable of performing. In addition, this document describes the user demographics that will be most likely to use the final product.

# Project Success

When completed, this product will be a highly accurate application of optical character recognition (OCR) software. The product will be able to read the text from an image file and translate that text into a computer-readable format, which can be modified or saved to a text file. Users will be able to take pictures of paper documents, input those pictures into this product, and receive text that is completely editable. The product will be dynamically trained using a series of complex algorithms, and this will give the product the capability to recognize different fonts, or even different languages.

This product will also contribute to the research and understanding of multiple fields, such as pattern recognition and machine learning. Most products that use OCR as their means of digital text translation do not reveal their algorithms or internal software designs to the public. On the contrary, the source code for this product will be open to the public, so it will serve as an excellent resource for students and software engineers who wish to learn more about the technology involved in OCR applications.

# Target Users

There are two separate target user groups for my project. These groups represent the two main purposes for the project, which are optical character recognition research and the development of a free and simple way to convert paper text into a digital format.

The first group is software engineers who are researching or implementing optical character recognition (OCR). The users in this group will benefit from my project by being able to analyze and dissect a working OCR system. These users will also be able to see the inner workings of my project, since I plan on releasing it as an open-source application.

The second group includes companies and organizations who need to convert already published books and documents into an electronic format. Many organizations, such as libraries, own large collections of paper books and documents that were published before the age of computers. As such, those organizations often need a way to convert their physical copies of published works into digital formats. Currently this conversion is only possible through the purchase of expensive commercial products or services that convert paper publications into electronic forms. That being said, those organizations in need of such a conversion would probably benefit the most from my project because my project would give them a free system to convert images with text into actual digital text. Of course, my project would not give the organizations everything they need, but it would be a useful tool to have available for organizations that want to perform their own conversions.

# Project Backlog

1. Previewing image before translation
   * Description: As a publisher, I want to preview the image I am about to translate so I can visually verify that I am translating the correct image.
   * Acceptance Criteria:
     1. The image previewer will display the image in a reasonable size.
     2. The previewer will provide an option to confirm that the image is correct, or to choose a new image.
2. Viewing translated text
   * Description: As a publisher, I need to view the text after translation so I can confirm that it has been translated correctly.
   * Acceptance Criteria:
     1. The translated text appears in a window automatically after the translation process is complete.
     2. If the text does not fit on the screen, the system will allow the user to scroll through the text.
3. Editing translated text
   * Description: As a publisher, I would like the ability to edit the text after translation, so that I can correct any mistakes made by the system or make revisions to the page.
   * Acceptance Criteria:
     1. The translated text can be edited at any time after translation in the same window in which it originally appeared.
     2. The standard editing commands (“cut”, “copy”, etc.) can be applied to the translated text.
4. Saving translated text
   * Description: As a publisher, I need to save the translated text so I can add it to an electronic publication at a later time.
   * Acceptance Criteria:
     1. The text-saving mechanism will provide a dialogue allowing the user to enter a filename and location for the new file.
     2. After selecting the option to save on the in the dialogue, the text will be written to a new file in the file system.
     3. Entering an invalid filename will return a message indicating the error.
5. Notification of translation failure
   * Description: As a publisher, I would like to be notified when an image cannot be translated so I can take steps to make the image more readable.
   * Acceptance Criteria:
     1. A message will appear automatically if a translation is not possible.
     2. The message will include details about the reason for which a translation could not be made.
6. Selecting different neural networks
   * Description: As an OCR engineer, I would like to choose between using the default neural network and implementing my own neural network, so I can study the differences between networks.
   * Acceptance Criteria:
     1. The system provides an option to use the default network or create a custom network.
7. Building custom neural networks
   * Description: As an OCR engineer, I need to specify the number of layers and neurons in my own neural network, so I can measure the efficiencies of various network structures.
   * Acceptance Criteria:
     1. The system provides the option to enter the total number of layers in a custom neural network.
     2. The system provides the option to enter the number of neurons in each layer of a custom network (excluding the input and output layers).
8. Training custom neural networks
   * Description: As an OCR engineer, I want the ability to train my own neural networks so I can use them in the application.
   * Acceptance Criteria:
     1. The system can train any feed-forward neural network given to it, so long as the input layer and output layer have the same structure.
     2. Any network trained by the system will be usable in OCR applications.

# Two Week Plan

1. Week One
   1. Translating images to text
      * Description: As a publisher, I need to translate image files into text so I can convert already published paper works into an electronic format.
      * Acceptance Criteria:
        1. The system will automatically translate the specified images into ASCII text.
        2. The translation will have an accuracy of at least 90 percent.
2. Week Two
   1. Translating images to text *(Continued from Week One)*
   2. Loading images
      * Description: As a publisher, I want to be able to load my own images of text pages into the application for translation, so I can convert any books into a digital format.
      * Acceptance Criteria:
        1. The image loader will provide a dialogue allowing the user to select a file.
        2. The loader dialogue will only allow image files to be selected.
        3. After file selection, the loader will close the dialogue.