Project: Document Layout Analysis

Opentext's Exstream software helps customers send enterprise scale, customized communications to their customers. One of the challenges that customers face is creating/importing new communications into Exstream. The goal of this project is to analyze an image of a communication and break that image into its' respective pieces.

This project will require reading, and possibly writing, various image file formats, analyzing the images to break up that image based on whitespace, outputting the results of the analysis in a computer readable format (probably JSON), and optionally providing a browser-based display of the original image with the analyzed zones highlighted. Opentext has an algorithm available for implementation, but students can develop an alternative algorithm if they would prefer.

The goal of this project is to analyze an image of a communication and break that image into its' respective pieces. The team would start with an image like shown in Figure 1. Base Image, and generate a set of coordinates that correspond to locations where the magenta and green lines intersect as shown in Figure 2. Analyzed Image.

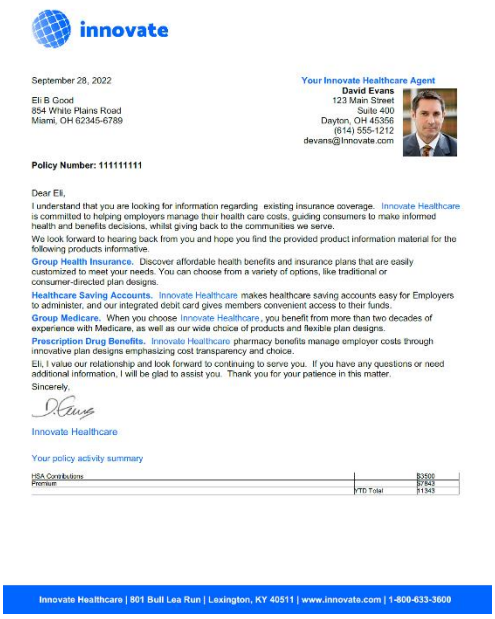

Figure 1. Base Image

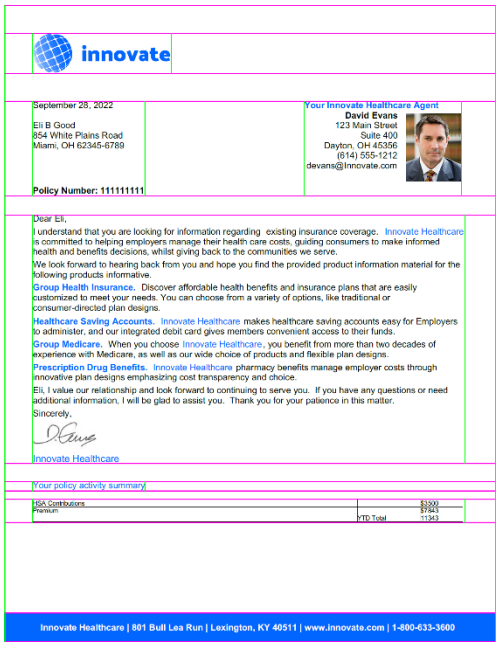

Figure 2. Analyzed Image

The customers for this project would be Nathan McConathy (OpenText, previous Senior design project customer), and Billy Kidwell (OpenText, previous Senior design project instructor).