EXPLORATORY RESEARCH

Food Bank
OF NORTHEAST GEORGIA

EXPLORATORY RESEARCH

Introduction

The Food Bank of Northeast Georgia is a non-profit organization dedicated to relieving hunger in 14 counties. With their headquarters in Athens, Georgia, the Food Bank of Northeast Georgia partners with around 200 organizations and agencies to distribute more than 13 million pounds of food to those in need each year.

The food bank operates as a 501(c)(3), which according to the Foundation Group, is a nonprofit that is allowed federal tax exemption. They gather donated items and purchased surplus food to process and store until they are ready for distribution to their various 501(c)(3) partner agencies, including soup kitchens, senior centers, food pantries, individuals, churches, and more. These partner agencies collect food from the food bank through their online inventory system or by selecting items from the shopping floor which operates similarly to a grocery store. While the food bank itself does not distribute any products, individuals and families in need can head to one of 200 locations to receive free-of-charge relief.

Before an individual or family can receive food from a partner agency or organization, they must fill out a paper copy of The Emergency Food Assistance Program (TEFAP) form which provides criteria for whether a household is eligible to receive food. The form has both English and Spanish options and requires an individual to fill out their personal information, the number of people in their household, their county, and a signature from the head of household to authorize the picking up of food from the distribution site. According to our client contact at the Food Bank of Northeast Georgia, Jenna Vaisvil, one new TEFAP form must be submitted at each distribution site that an individual visits. While filling out these TEFAP forms is a requirement by the USDA for all food banks, the information collected by the Food Bank of Northeast Georgia's partner agencies is not verified. Vaisvil states this is so no one feels discouraged to receive aid if they feel they need it.

At the moment, the collected TEFAP forms at the Food Bank of Northeast Georgia's partner agencies are stored at each facility in case the USDA audits the food bank. Currently, no data is being pulled from the TEFAP forms, but Vaisvil hopes that through the digitization of the forms, the food bank can collect useful information about county breakdowns. This would allow the food bank to more effectively reach those in need by understanding which areas in Northeast Georgia are most vulnerable.

Sources & Products

After doing more research on the Food Bank of Northeast Georgia and the problem at hand, we looked into some of our options for digitizing and using the TEFAP forms effectively. This started with understanding more about how forms are digitized. Through OCR (Optical Character Recognition) technology, forms are scanned and read by computers to put their information onto a computer screen.

According to a Youtube video created by Techquickie, OCR technology works by removing the background to focus on the text. It then converts the text to only black and white, aligns the text in a straight line, and compares each letter to a known font database. OCR software can also utilize a dictionary as a reference to be more accurate in word creation. Once we discovered that this technology would help with our first step in transferring our TEFAP forms to a computer, we set out to look for an OCR service that would best suit our needs. This definitely came with its challenges, as there are multiple softwares out there with a vast array of features.

Considerations for the best software include: price, easy to use, free option/ test run, able to read handwriting, accuracy, future costs, data extraction (not solely digitizing the form), and compatible devices. We considered and extracted information from second party reviews including Techradar, Brian Kent of Crosstab Kite, Capterra, and of course, John Weatherford. One major roadblock we hit was understanding the most efficient process to link OCR to a database. According to SimpleSoftware, one solution to our problem was separating the two steps completely with two different softwares. A more efficient option mentioned by SimpleSoftware was automated packages that combine these two steps. Below are the different products we researched, features, what we liked, and what we didn't like. With these considerations in mind, we picked Google Document AI to test out within our alpha project. By testing out these softwares, we are providing our client with the service that best fits their needs and makes our work easier in the future.

1. Abbyy

Abbyy is a Digital Intelligence company that provides numerous services to global companies. Their products and solutions are vast, but "FlexiCapture" is our focus here. With FlexiCapture, businesses can transform documents into data with advanced technology. Flexicapture "brings together the best NLP, machine learning, and advanced recognition capabilities into a single, enterprise-scale platform to handle every type of document, from simple forms to complex free-form

documents, and every job size, from ad hoc single documents to large batch jobs requiring tough SLAs."

Abbyy seems to be a very complex service for more detailed projects, so it might be better to take a simpler route. We have requested a demo for this service and will have to wait to hear back from them. FlexiCapture pricing can range from $29.99 to $299.99 per month based on the number of pages. We are not able to test this service out before checkpoint 1 because we have not heard back from a sales associate about getting a trial.

2. Amazon Textract

This AWS software stood out to us in the Brain Kent review due to its ease of use and price. With Amazon Textract, you can "read and process any type of document, accurately extracting text, handwriting, tables, and other data…and act on the information extracted…". The AWS free tier is worth a try as it lasts one year and allows you to try out multiple services. According to the Amazon Textract webpage, it is covered in the Free Tier. One negative found in the review mentioned that it requires asytronchus PDF processing, therefore is slower to use than its competitors. Pricing is a pain point for this service, as it charges by region and combination of PDF elements.

3. DocuClipper

This service extracts data from PDF documents and imports it into Excel. We can customize the PDF to tell the software what to read, which makes this service seem simple to use. With a free 14 day trial, this service definitely seems like a top contender to put to the test. The only hesitation we have is planning out the steps we would have to take after converting the PDFs to an Excel sheet, and if this is the most time efficient method. Pricing wise, we assume we would have to commit to the $59/month which is 500 pages or $29/month for 200 pages depending on what works best for us. After testing out this service, we conclude that it lacks accuracy and ease of use. The trial we started was very glitchy,

3. Docparser Software

As a fully automated "scan to database" software, Docparser seems to be more effective than manual extraction. Docparser can "automatically extract data from scanned documents and transfer to the database for you at a fraction of the costs."

First you must scan the documents, docparser captures the data on the documents and moves your data into your SQL or NoSQL database. There are three ways to upload your data to a database depending on the specific one in use. After testing out this service we were surprised to find out that the software does not read handwriting and is therefore not going to be an option for us.

4. Google Document AI

Google's Document AI service has a lot of processors under its software that we were excited to try out. According to Brain Kent, it's form parser software was the most accurate when compared to other large services. It's OCR service is able to "identify and extract text in different types of documents.This processor allows you to identify and extract text from documents in over 200 languages for printed text and 50 languages for handwritten text" After using the "Try It" feature, we were really impressed with the ease of use and the accuracy. Out of all of our tests and trials, this one gave us the best results and therefore, we are going forward with this service and using it within our alpha test to show the class more about OCR and what we've learned.

Further Research

We were pretty set on using an online software service, but wanted to also consider iOS options. HiTechNectar's list of OCR Mobile Apps was a good place to start to look into the top OCR mobile applications. The problem we faced with apps was trying to find something that could handle the scale and structure we were looking for. CamScanner was successful in translating our documents into full text, but didn't allow us to select certain areas to pull data from. It also didn't allow us to create a template like the online services allowed. It was also hard to navigate when the service wanted us to pay for most of the elements. SmartLens was another really cool app we looked at. It's application was more for fun and casual use, scanning and identifying text, objects, and even celebrities. It was interesting but not what we are looking for. Text Scanner looked hopeful because it is the "best bet if you want an app that supports handwriting recognition." Text Scanner was interesting because we could select the text we wanted to read, however we couldn't create a template and it was not very accurate. This app was probably more of a one time deal. We tried a few other iOS versions but concluded that online softwares is better to use long term. They are more accurate, easier to

test out with trials, and more likely to connect to our database that we are planning to work with in the future.

After digitizing our forms, we question how we will be able to utilize the data. In our meeting, Jenna Vaisvil pointed out the need for data visualization using both the county and household size to paint a picture for the organization. With their new SNAP program, mapping out neighborhoods that need assistance would be a great step. According to a case done by McKinsey & Company, data has been vital for certain food banks during the pandemic. Specifically, Second Harvest Heartland "partnered with McKinsey to develop a demand model to forecast the supply that it would need, based on various scenarios of future demand." Their data visuals include best and worst case scenarios of food demand and total food supply. The data and visualizations they are working with may be different than our project's needs, but are still inspirations for creating change and decisions with food bank data. Not only do we want to complete detailed data analysis for the FBNEGA, but we also want to make