

Working Prototype Known Problems Report

3/3/21

List all problems with your software that your team has identified. Include in the item entry the input/action that causes failure; location of fault (if known); possible action for removal of fault.

- Issue: OCR Parsing results are often inaccurate. Some receipts will not have their items parsed out, and some will register additional fees such as sales tax as line items.
 - Replication of error: Unclear photos or photos of receipts with wrinkles/smudged ink will often not produce a good ocr result. Receipts without dollar signs in front of prices seem problematic
 - Location of Fault: As we are using a 3rd party receipt parsing API by Cloudmersive, the fault lies within the OCR library and parsing algorithm used by the company.
 - Possible Fix: There are two routes to take here, both involve changing our OCR/Parsing component. One option would be to select another receipt parsing API/Library. We could not feasibly do this because most enterprise level, good quality parsing libraries cost upwards of \$400 a month to use. We could also create and train our own ML model to attempt creating a more accurate parser. However, that requires additional computing resources which is typically costly depending on the scale of training.
- Issue: OCR Parsing takes a long time
 - Replication of error: All receipt images, especially long receipts or large images can take upwards of 20 seconds to process
 - Location of Fault: (Same as above issue)
 - Possible Fix: (Same as above issue)