

Document Processing

Referral form extraction & Output form inserts

31/08/2024

The program currently extracts by searching for everything in between two sections.

For example, it will automatically look for “Section 1” and “Section 2” as a start and end point, if the program successfully finds those fields, it will proceed to extract all the information in between the two sections. This is a search and extract, as the program parses through the information, will look for key words like “Name (First & Last), and extract the information next to it “John Doe”.

This works the same way for the assessment form itself, the program searches for a start and end point, and inserts the data in between the fields, with a search and replace.

For example, it will look to find “Client details” and “Disability / Diagnosis Details” for a start and end point, and then parse through the information between them, looking for key words “Last name”, “First name”. and then basically rebuild that information while inserting the correct information. “Last name: Doe”.

All information is currently re-written with the format already provided in the document. So that this point, all information is in bold format. But as this is a starting prototype/proof of concept to prove that it can be done. Should be find at this moment.

The program currently excels at extracting all available data from the referral form, specifically for the client details section of the assessment information form. We have successfully automated the processing of client details, disability/diagnosis, and alternative contact information.

However, as discussed on Friday, the remaining sections of the assessment information form are dependent on further steps in the process chain. There are challenges with incomplete or poorly labelled information on the referral form, making it difficult to accurately determine the placement of other data on the assessment form.

Section 1 of the document has been successfully extracted, but there are fields on the assessment information form that are still blank due to not being on the referral.

- Type of residence
- Relationship to others in the household
- Preferred language
- Community services card
- CSC expiry date
- Home phone
- Work phone

Section 2, all data except for the attach confirmation of diagnosis form has been extracted.

Section 3, of the referral form, currently takes the disability name/type information, and inserts all name/types into the Disability / Diagnosis field.

Currently this only correctly works when properly comma separated.

The rest of section 3 will need to be clearer on the assessment form itself, on where that information goes.

Section 4, I am currently assuming, goes on the Other Information section of the assessment form.

Section 5, alternative contact works perfectly, but the referral form is missing a few fields.

- Address
- Home phone
- Work phone
- Date of birth

Section 6 Referrer's contact details, does not currently exist on the assessment form, so there is no where to put it in.

Section 7 only a few pieces of information of the consent and privacy are able to be lined up with the assessment form.

Section 8, also currently has no place to insert the data to the assessment form.

For the assessment form itself, apart from the findings listed above. Everything including and below the Consent for information sections do not exist on the referral form, but as stated above, will come from the second step of the process.

At current, as the clients family fills out the referral form, there are a lot of edge cases for human error when filling out the form, which will result in errors in the assessment form. As we can adjust the referral form, I believe it, some helpful information as to how to properly fill the fields out would help reduce the errors.

As for the assessment form. If there is a possibility of writing the fields in a certain way, if a search and replace approach of data insertion is fine, would help reduce a lot of the workload on the program. For example

Client Details

- **Last name:** [client_last_name]
- **First name:** [client_first_name]
- **NHI number:** [client_nhi_number]

As the program currently parses through the document to look for start and end points, and then checks through the information between to find the correct field. I feel using a [information] approach would help preserve formatting, and remove a lot of document rebuilding.