**Different Parsing Comparison->**

**Objective->**

The objective is to identify the different Parsing Techniques like Default Parser, FM parser, Textract and Data Automation parsing along with the suitable uses cases where to use which Parsing techniques.

**Approach->**

We are selecting different types of scenarios mentioned below to check the performance of different Parsing Techniques

1. Normal Text
2. Flowchart Data
3. Complex Table Data: (It is having the icons data in one of the columns)
4. Analyzing Minute Details in Images
5. Text Data in Tables
6. Complex Images with Text

**Parsing Techniques Used->**

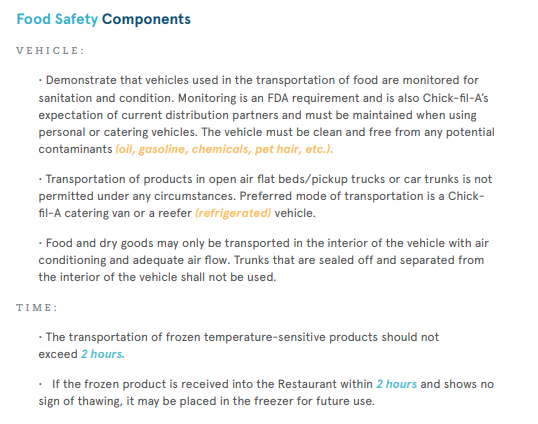
1. Default Parser
2. AWS Textract
3. Foundation Model as Parser
4. Data Automation as a Parser

**Normal Text->**

**Test Data->**

A white text on a white background

Description automatically generated

****

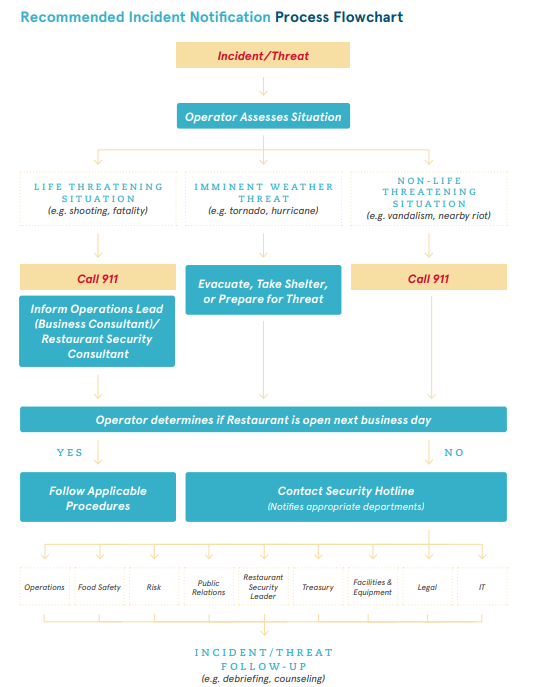
**Results Comparison->**



**Conclusion->** The results generated by Data automation parser are better, however we can also use Default parser as it is free of cost.

**Flowchart Data->**

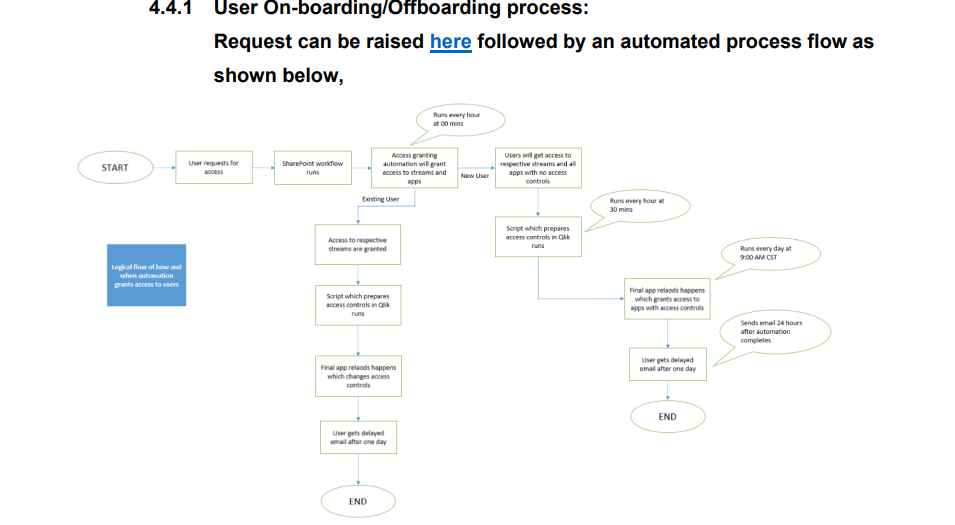
**Test Data-> Example 1**



**Results Comparison Experiment 1->**



**Example 2->**

****

**Results Comparison Experiment 2->**



**Conclusion->** The results generated by the Data automation is better than other parsing techniques, and hence should be used while dealing with the Flow chart type of data.

**Complex Table Data: (It is having the icons data in one of the columns)->**

**Test Data->**

**A screenshot of a computer security

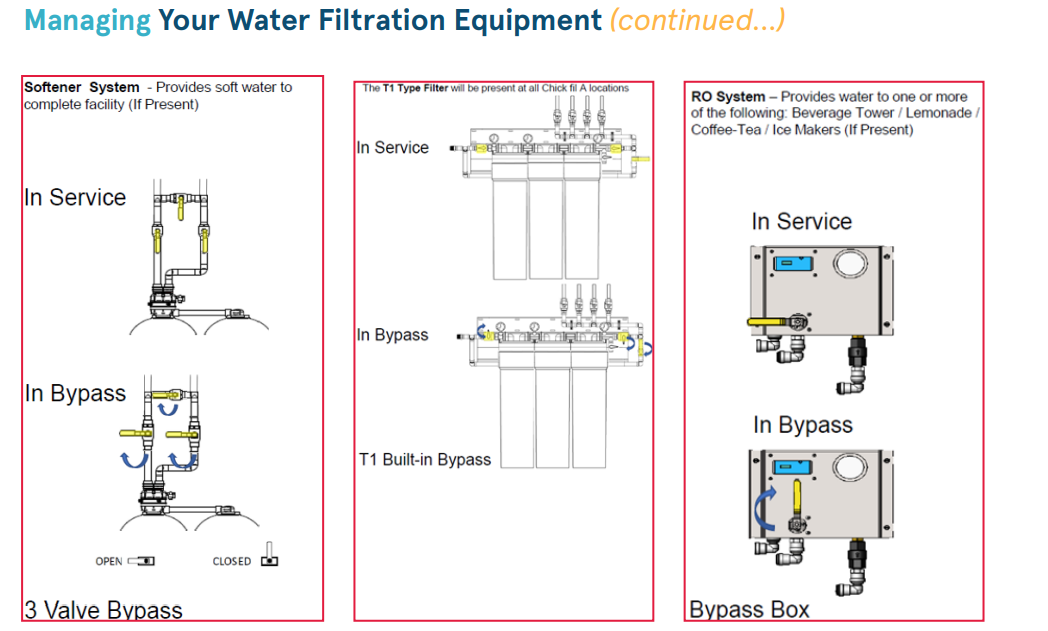
Description automatically generated**

**Results Comparison->**

**Conclusion->** The Data automation Parser was able to detect the Images present in the Table and was able to generate the proper result, hence it should be used for table data with images.

**Analyzing Minute Details in Images->**

**Test Data Experiment 1->**

****

**Results Comparison Experiment 1->**



**Test Data Experiment 2->**

**A diagram of a jumper wire

Description automatically generated**

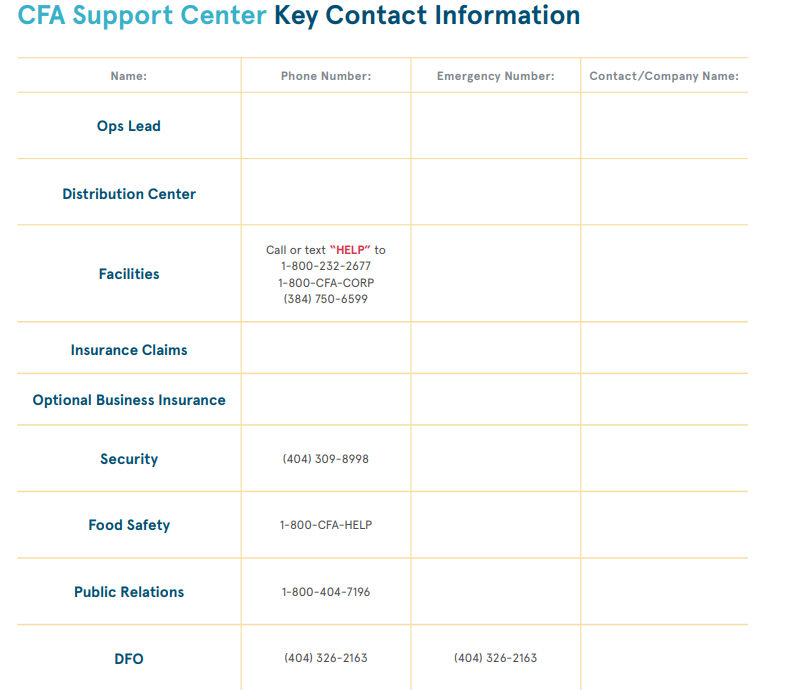
**Results Comparison Experiment 2->**



**Conclusion->** None of the parsing technique was able to give the accurate result but the results generated by Data Automation is better as compared to others, hence should be preferred for Data with complex images.

**Text Data in Tables->**

**Test Data Experiment 1->**

****

**Results Comparison Experiment 1->**



**Results Comparison Experiment 2->**



**Conclusion->** All the techniques performed well; Default parser can be used since it is free of cost.

**Complex Images with Text->**

**Test Data Experiment->**

**A diagram of a standard front label size

Description automatically generated**

**Results Comparison->**



**Conclusion->** The results generated by Data Automation was better than other techniques, however it also does not give the whole answer.

**Final Conclusion->**



**Evaluation Data->** [**Data**](Data)

**Detailed Analysis is available here->** <Parsing_Technique_Evaluation.xlsx>

**Cost Analysis->**



**Conclusion->**

Data Automation can perform better in almost all the scenarios, however there is cost involved to it.

We can use Default parser when dealing with Normal text and Text in Table as there is no cost involved in it.