

# Hoja de Sesión

**Tester:** Yismary Quintero

**Fecha:** 22/11/2023

**Hora de Inicio:** 19:00hs

**Duración:** 45m-2hs

## Desglose de Tareas

**Set up de la sesión:** 10%

**Diseño y ejecución:** 80%

**Reporte e investigación:** 10%

## Misión

Validar que la data extraída y transformada (documento PDF a JSON) cumpla con consistencia, completitud, contextualización e integridad de datos.

## Áreas

Funcionalidad: Fase procesamiento de datos

Estrategia: Testing Exploratorio

## Archivo de datos

instruction\_handbook-Copiar.pdf

texto\_extraido\_modificado3.pdf

texto\_extraido\_modificado3.json

## Notas de prueba

Convertimos el archivo json en formato pdf

Comparamos **instruction\_handbook.pdf** vs **texto\_extraido\_modificado3.pdf**

## BUG-01 Omisión de palabras cortas en archivo formato JSON

**Descripción:** Se ha detectado un problema consistente en la omisión de palabras cortas en el texto resultante. Las palabras omitidas son generalmente preposiciones, artículos y conjunciones, incluyendo pero no limitándose a: "and", "of", "the", "for", "as", "is", "on", "under", "out", "all", "that", "to", "with", "any", "by", "in", "have", "at", "each", "those". Estas omisiones afectan la integridad y la comprensibilidad del texto extraído.

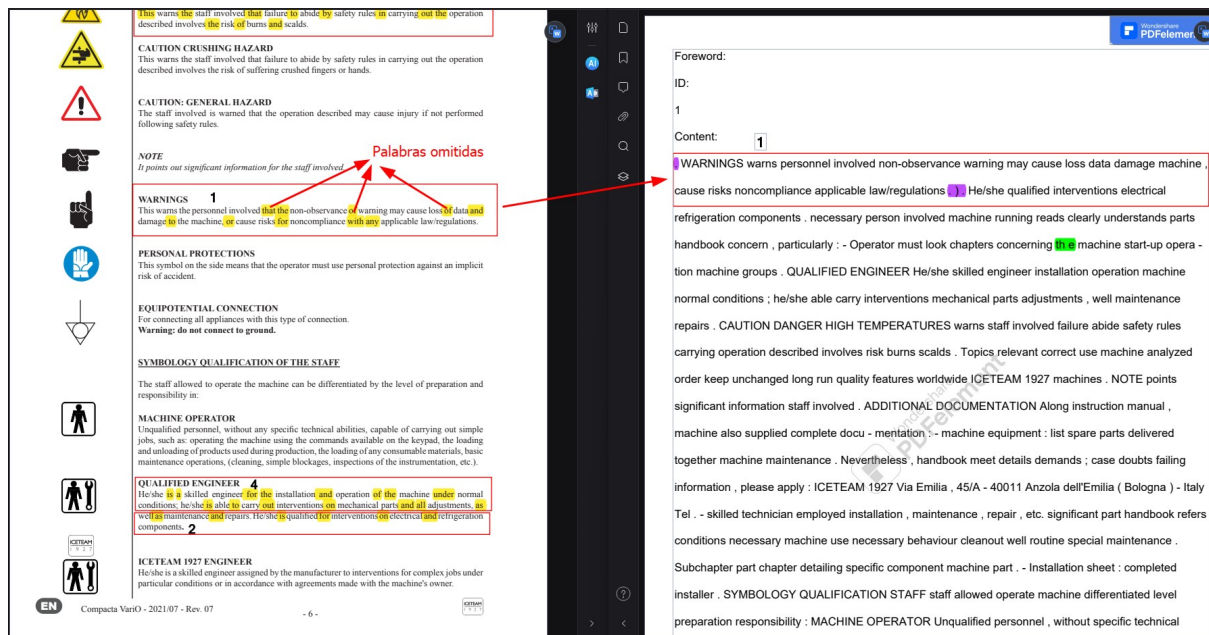
## Severidad Media

## Prioridad Alta

**Resultado Esperado:**El archivo JSON debería contener una transcripción fiel y completa del texto original del PDF, incluyendo todas las palabras, independientemente de su longitud.

**Resultado Actual:** El archivo JSON generado omite sistemáticamente palabras cortas específicas, lo que resulta en frases incompletas y posiblemente altera el significado o la comprensión del texto.

**Evidencia:**



## BUG-02 Inconsistencia en la Secuencia del Texto dentro de parrafos en Json

**Descripción:** La secuencia del texto en el archivo JSON no coincide con la del documento PDF original, presentando un orden diferente en las frases y párrafos. Este desorden podría tener implicaciones graves en el entrenamiento de modelos de machine learning para un chatbot, ya que la secuencia de datos es esencial para el aprendizaje contextual y la generación de respuestas coherentes.

**Severidad Alta**  
**Prioridad Alta**

**Resultado Esperado:** El texto en el archivo JSON debe reflejar con precisión la secuencia y el orden del contenido del documento PDF original.

**Resultado Actual:** El archivo JSON muestra un orden de texto diferente al del PDF original, alterando significativamente la estructura y la coherencia de la información.

### Evidencia:

The image shows a side-by-side comparison of a PDF document (left) and a JSON file (right). Red arrows indicate mismatches in the order of text segments between the two formats.

**PDF Document (Left):**

- CAUTION CRUSHING HAZARD**: This warns the staff involved that failure to abide by safety rules in carrying out the operation described involves the risk of suffering crushed fingers or hands.
- CAUTION: GENERAL HAZARD**: The staff involved is warned that the operation described may cause injury if not performed following safety rules.
- NOTE**: It points out significant information for the staff involved.
- WARNINGS 1**: This warns the personnel involved that the non-observance of warning may cause loss of data and damage to the machine, or cause risks for noncompliance with any applicable law/regulations.
- PERSONAL PROTECTIONS**: This symbol on the side means that the operator must use personal protection against an implicit risk of accident.
- EQUIPOTENTIAL CONNECTION**: For connecting all appliances with this type of connection. Warning: do not connect to ground.
- SYMBOLGY QUALIFICATION OF THE STAFF**: The staff allowed to operate the machine can be differentiated by the level of preparation and responsibility in:
- MACHINE OPERATOR**: Unqualified personnel, without any specific technical abilities, capable of carrying out simple jobs, such as: operating the machine using the commands available on the keypad, the loading and unloading of products used during production, the loading of any consumable materials, basic maintenance operations, (cleaning, simple blockages, inspections of the instrumentation, etc.).
- QUALIFIED ENGINEER 4**: He/she is a skilled engineer for the installation and operation of the machine under normal conditions, he/she is able to carry out interventions on mechanical parts and all adjustments, as well as maintenance and repairs. He/she is qualified for interventions on electrical and refrigeration components.
- ICETEM 1927 ENGINEER**: He/she is a skilled engineer assigned by the manufacturer to interventions for complex jobs under particular conditions or in accordance with agreements made with the machine's owner.

**JSON File (Right):**

```
Foreword:
ID:
1
Content:
. WARNINGS warns personnel involved non-observance warning may cause loss data damage machine ,
cause risks noncompliance applicable law/regulations . 2: He/she qualified interventions electrical
refrigeration components . necessary person involved machine running reads clearly understands parts
handbook concern , particularly : - Operator must look chapters concerning the machine start-up opera -
tion machine groups . QUALIFIED ENGINEER He/she skilled engineer installation operation machine
normal conditions ; he/she able carry interventions mechanical parts adjustments , well maintenance
repairs . CAUTION DANGER HIGH TEMPERATURES warns staff involved failure abide safety rules
carrying operation described involves risk burns scalds . Topics relevant correct use machine analyzed
order keep unchanged long run quality features worldwide ICETEM 1927 machines . NOTE points
significant information staff involved . ADDITIONAL DOCUMENTATION Along instruction manual ,
machine also supplied complete docu - mentation : - machine equipment : list spare parts delivered
together machine maintenance . Nevertheless , handbook meet details demands ; case doubts failing
information , please apply : ICETEM 1927 Via Emilia , 45/A - 40011 Anzola dell'Emilia ( Bologna ) - Italy
Tel . - skilled technician employed installation , maintenance , repair , etc. significant part handbook refers
conditions necessary machine use necessary behaviour cleanout well routine special maintenance .
Subchapter part chapter detailing specific component machine part . - Installation sheet : completed
installer . SYMBOLGY QUALIFICATION STAFF staff allowed operate machine differentiated level
preparation responsibility : MACHINE OPERATOR Unqualified personnel , without specific technical
```

Red arrows point from the JSON text to the corresponding (or mismatching) text in the PDF, showing that the sequence in the JSON does not match the original document's structure.

### BUG-03 Inconsistencia en la secuencia de informacion en Json

**Descripción:** El archivo JSON debe reflejar la secuencia del contenido del PDF original con exactitud, manteniendo la organización lógica y estructural del texto.

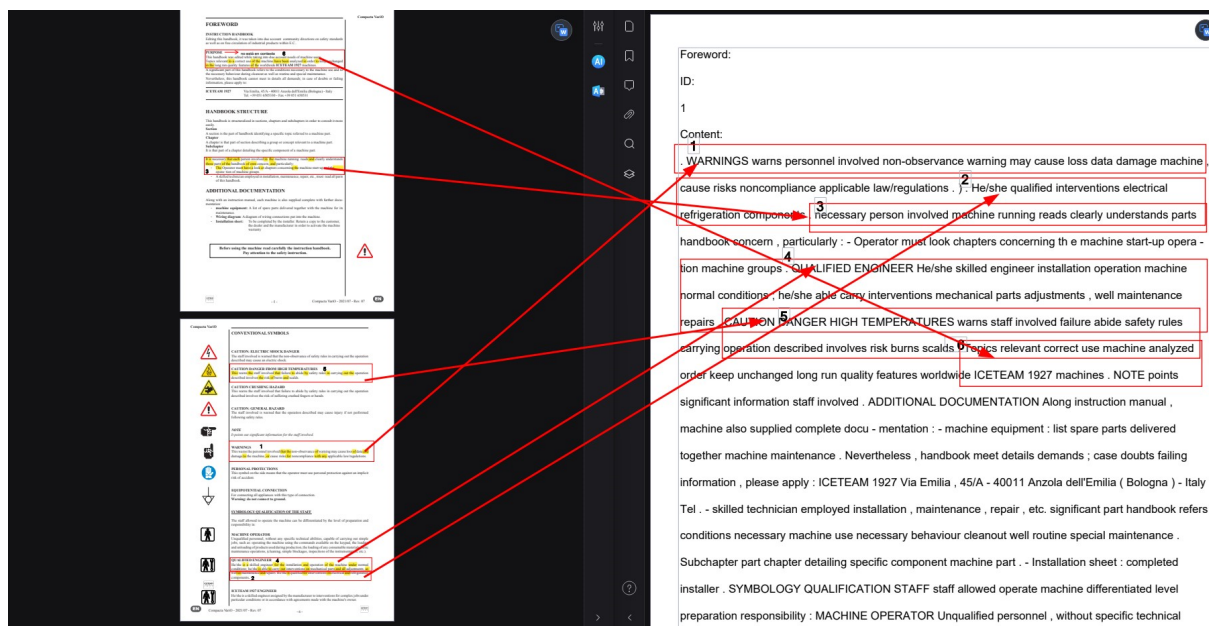
**Severidad Alta**

**Prioridad Alta**

**Resultado Esperado:** El archivo JSON debe reflejar la secuencia del contenido del PDF original con exactitud, manteniendo la organización lógica y estructural del texto.

**Resultado Actual:** Hay una discrepancia evidente en el orden del contenido en el archivo JSON, con frases y segmentos de texto que aparecen desordenados en relación con el documento original.

#### Evidencia:



## BUG-05 Pérdida de contexto en la informacion en Json

**Descripción:** Se ha identificado que segmentos de texto extraídos y convertidos en JSON no mantienen las referencias de títulos y subtítulos presentes en el documento PDF original, lo que es crucial para la comprensión contextual de la información.

**Severidad Alta**  
**Prioridad Alta**

**Resultado Esperado:** El archivo JSON debe incluir la estructura de títulos y subtítulos del documento PDF original, proporcionando el contexto necesario para cada segmento de texto.

**Resultado Actual:** Los datos en el archivo JSON aparecen sin las referencias contextuales de títulos y subtítulos, resultando en contenido que parece carecer de contexto.

## Evidencia

