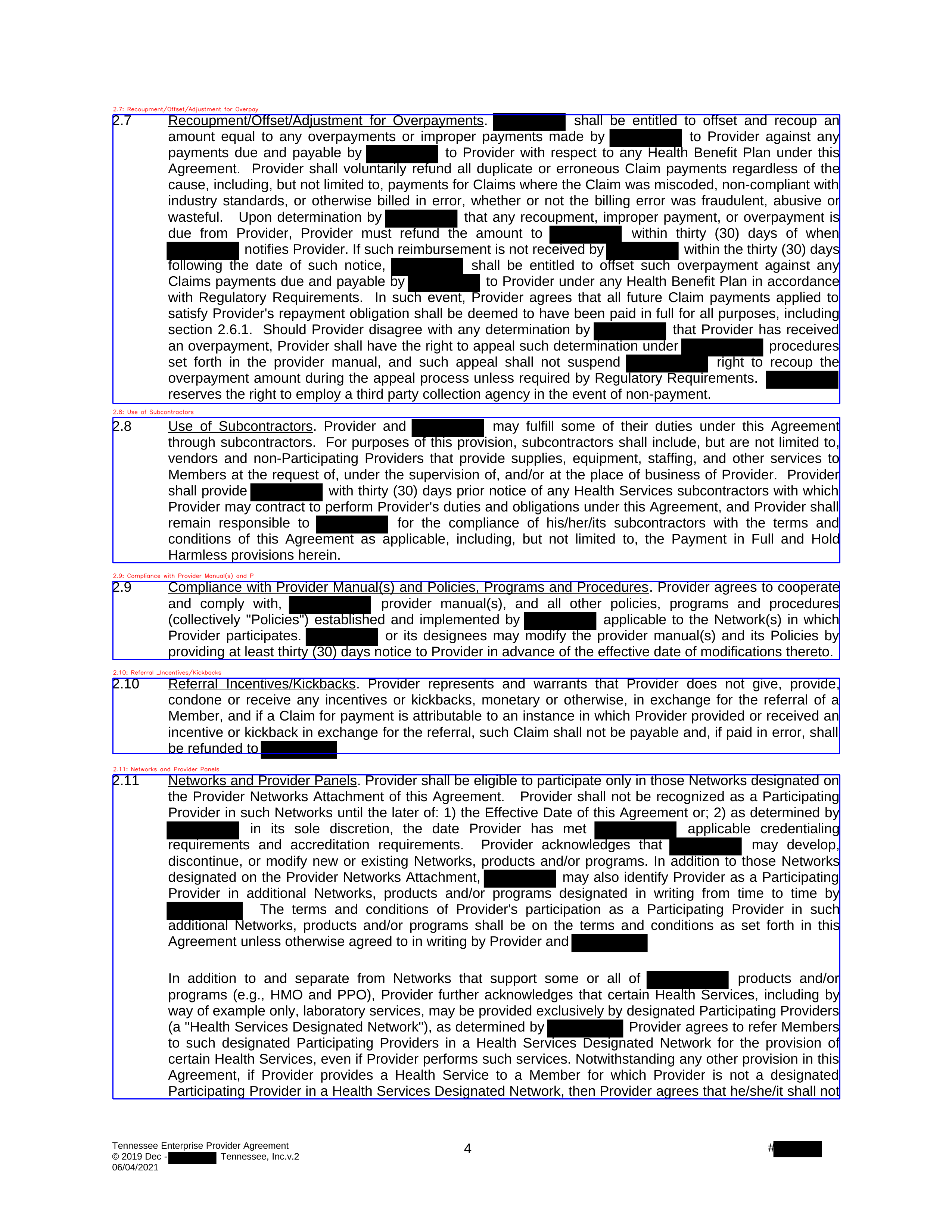
**EXTRACTION OF CLAUSES FOR 1,2,3 ATTRIBUTES:**

1. Input pdf -> Section wise labelling and Identification using Tesseract OCR (hOCR).



1. Searching for the required section in the structure of the pdf generated bby specific keywords for different attributes

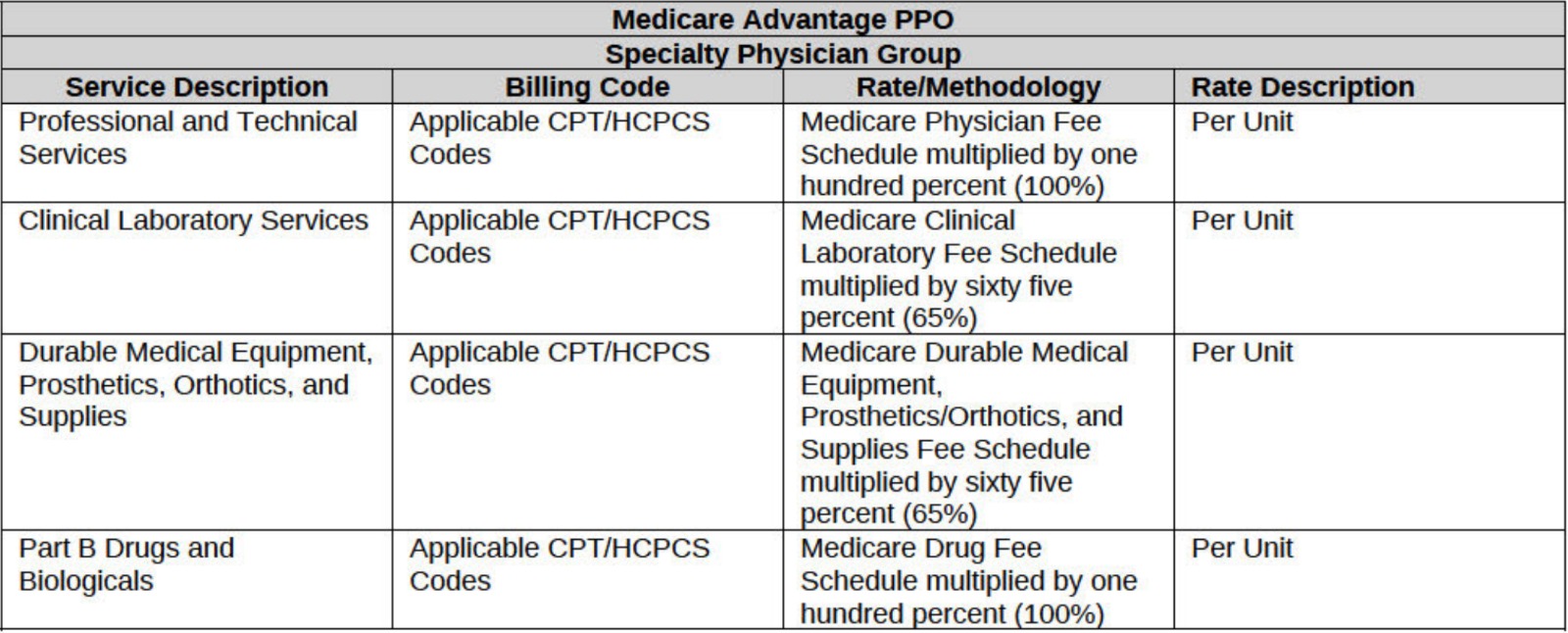
* "Attr1": ["Submission and Adjudication of Medicaid Claims"],
* "Attr2": ["Submission and Adjudication of Medicare Advantage Claims"],
* "Attr3": ["Networks and Provider Panels"]

1. Identified and cropped the corresponding sections. And extracted text from them using tesseract OCR.

**EXTRACTION OF CLAUSES FOR 4,5 ATTRIBUTES:**

1. Identifying key page numbers by searching for the keyword ‘Service Description’.
2. Extracted Tables using Contour detection in OpenCV

Example:



1. Extracted the text using image to csv libraries.

Example:  
