Q3: DESIGN A SYSTEM THAT CAN EXTRACT CONTEXTUAL CHUNKS OF TEXT FROM A GIVEN PDF. Instructions: • Refer sample.pdf for sample input & output. Text chunks highlighted in YELLOW box is the expected output. • Refer output.pdf for a sample output. • No coding is required. Design only the data pipeline and technology / libraries that would be required to build the system.

PDF Reader

Raw Text

Sentence segmentation

Tokenization

POS tagging

Entity Detection

Relationship Extraction

Rule creation