

FEDERAL INSTITUTE OF SCIENCE AND TECHNOLOGY (FISAT)



Hormis Nagar, Mookkannoor PO, Angamaly, Kochi
Accredited by NAAC with 'A+' Grade

DEPARTMENT OF COMPUTER APPLICATIONS

SYNOPSIS OF THE MAIN PROJECT

Name of the Student	Nimi Harison
Batch, Roll Number	MCA B,39
Contact Number & Email id	nimiharison2002@gmail.com , 9746812588
Name of Project Guide	Dr.Deepa Mary Mathews
Date of Submission	15-02-2024
Project Title	Document Parser
Area of the Project	Natural Language Processing
GitHub ID	nimi202
<p>Description of Project:</p> <p>The Document Parser stands as a robust tool that combines advanced Natural Language Processing (NLP) techniques with custom rule-based data extraction methods. Employing NLP algorithms, it comprehensively analyzes diverse documents, offering capabilities such as part-of-speech tagging and named entity recognition. Its custom rule-based approach ensures adaptability to various document formats, allowing the tool to accurately extract structured data from unstructured text. Beyond keyword matching, the Document Parser understands contextual nuances, providing a nuanced understanding of information. With dynamic document type recognition and a user-friendly interface, users can efficiently configure extraction rules without extensive coding. The tool's scalability, real-time feedback, and potential integration options make it a versatile solution for automating data extraction from documents, offering efficiency and accuracy in information retrieval.</p>	
Front End & Back End Tools	React & Python, Java, MySql