User Manual

Student 1 – Lorena Gomez – 21734359

Student 2 – Darragh Manning - 21506373

Supervisor - Tomas Ward

Project Name – TimelineXtract

Last edited on: 02/05/2025

Abstract - TimelineXtract is a system designed to automate the extraction of questionnaire schedules from complex clinical trial protocols. Clinical trial PDFs are often lengthy, inconsistently structured, and difficult to navigate, making manual data extraction slow and error prone. Our solution combines Adobe PDF Extract API for precise document structure recognition, GPT-4 for semantic data extraction, Django and React.js for processing and presentation, and MongoDB for secure storage.

The project overcame major challenges such as inconsistent table formats, varying questionnaire names, and aligning data from different extraction sources. Through extensive prompt engineering, vector similarity matching, and real world testing with protocols from ClinicalTrials.gov, TimelineXtract delivers clean, structured outputs that streamline clinical research.
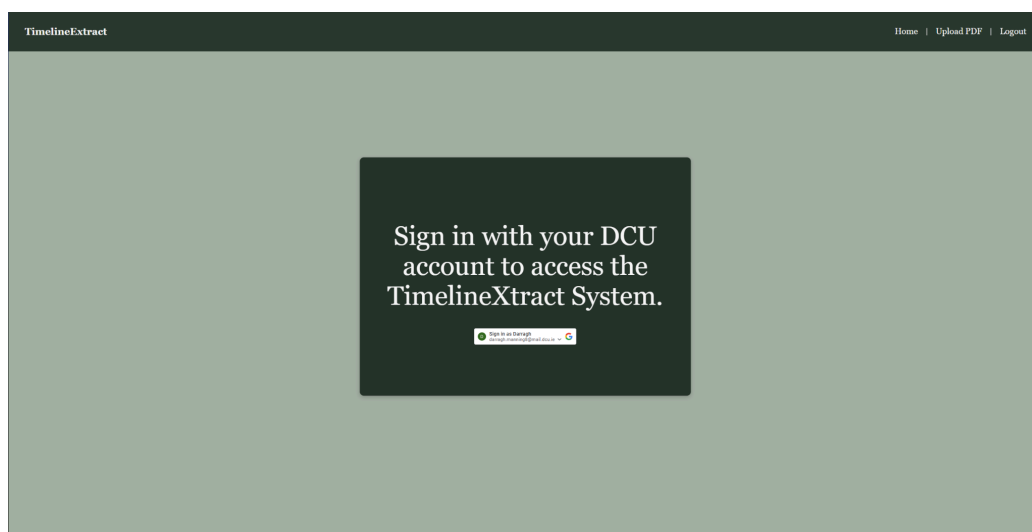
Introduction - TimelineXtract is an application that allows you to upload a Clinical Trial Protocol in a PDF and extract the relevant timeline information from that file. It is targeted at patient participants in a clinical trial study. This document details how to use the application.

System Requirements - The application uses react and so is supported natively on Google Chrome, Firefox, Microsoft Edge and Safari. [1] It is compatible with desktop, mobile and tablet.
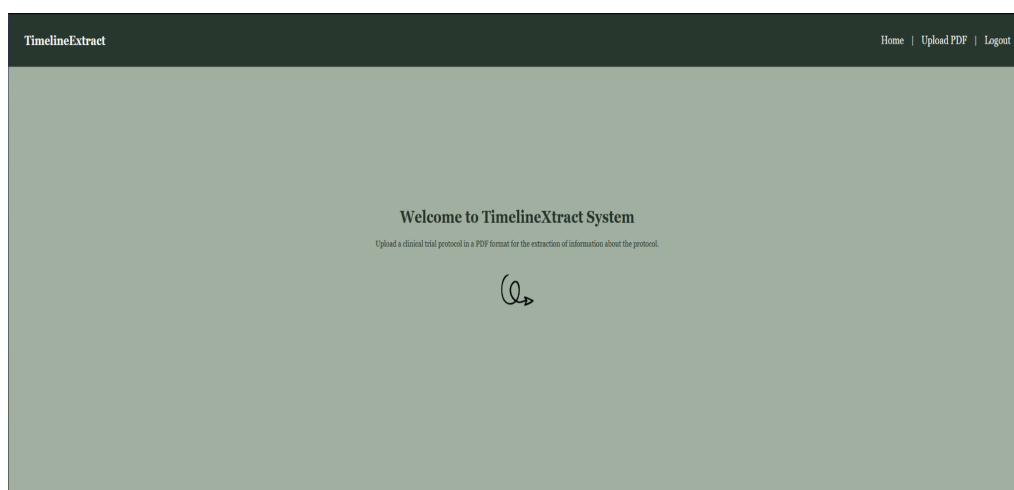
Login Page - The project is deployed with Render (Backend) and Vercel (Frontend) and can be accessed at this link https://2025-csc1097-mannind-8-gomezl-3.vercel.app/

Authentication - You are able to login with a dcu email address only, a personal email will not work.
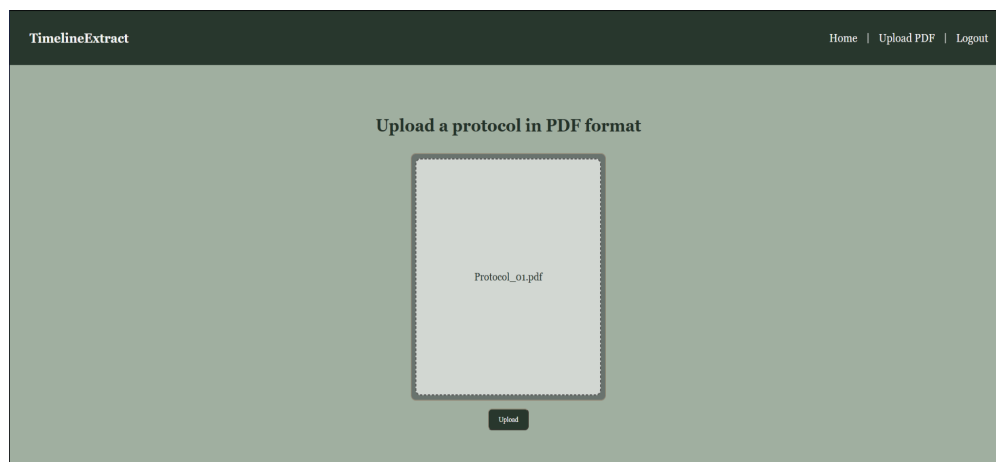
Login Page:

Home Page:

Features - The main feature of TimelineXtract is to upload a pdf. Simply click 'Upload PDF' on the navigation bar to open the upload page.
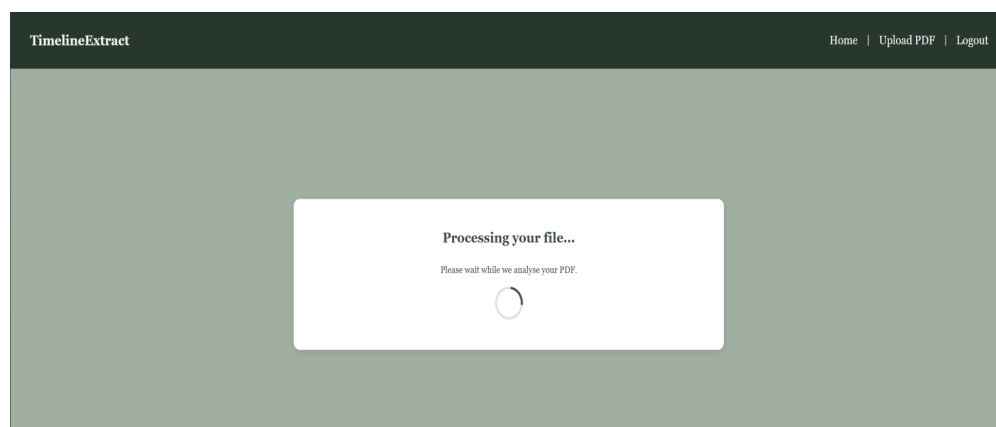
Upload Page:

Drag and drop a file into the box or click it to open File explorer. Select a PDF containing clinical trial protocol.

It will take some time to process the file.

Results - Once the file has been processed you should get a response that looks something like this.

Results Page:

Use the arrows to expand or collapse the data that is most relevant to you.

Limitations

- The application runs best on desktop.
- The requests to the backend can 'wind down' with inactivity meaning it could take a long time for the first request to the backend to go through. Afterwards they will run like normal (near-instant assuming a good internet connection).
- The application is using only the free option for the deployment platform and is not handling more than a small amount of traffic.
- It may take a few minutes for the system to generate the results.

References

1. JavaScript Environment Requirements – React'. Accessed: Apr. 30, 2025. [Online]. Available: https://legacy.reactjs.org/docs/javascript-environment-requirements.html