School of Computing

TimelineXtract

Student: Lorena Gomez & Darragh Manning, BSc in Computer Science
Supervisor: Professor Tomas Ward

Introduction

- TimelineXtract is an application that speeds up and simplifies the process of participating in a clinical trial by extracting the relevant questionnaires and timeline events from a clinical trial protocol.

- A participant of a clinical trial can upload a PDF of their Clinical Trial Protocol and TimelineXtract will respond with the relevant information.

Methods

- The clinical trial protocols were taken from a large public database of trials maintained by the United States government.

- GPT-4 extracts the information using our carefully engineered prompt. Adobe API then extracts the tables and our system filters for the desired information.

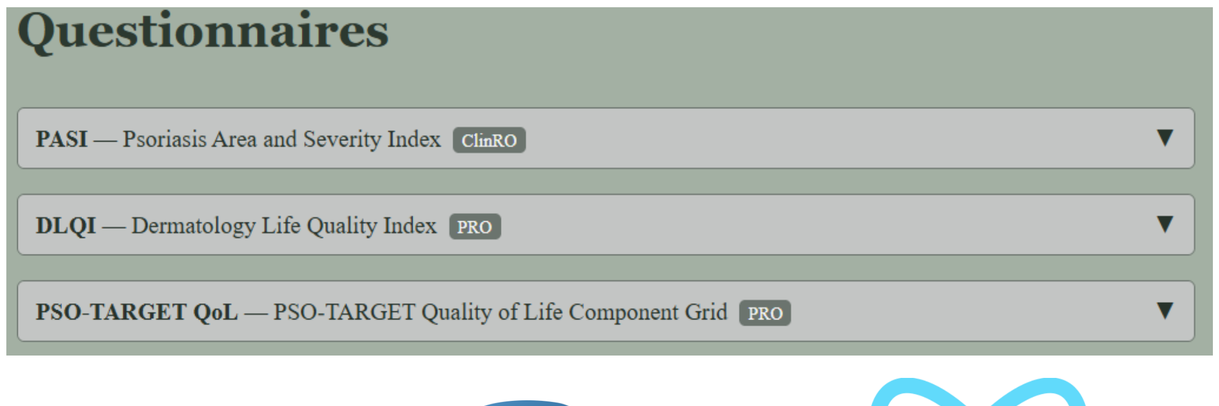
- Timeline Xtract displays the information to the user in a clear and organised way.

Benefits

- Saves time and reduces manual labour of clinical trial participant.

- Reduces chance for participant error, speeding up the trial and ensuring potentially life saving medication can make it to the market sooner.

- TimelineXtract uses secure authentication.

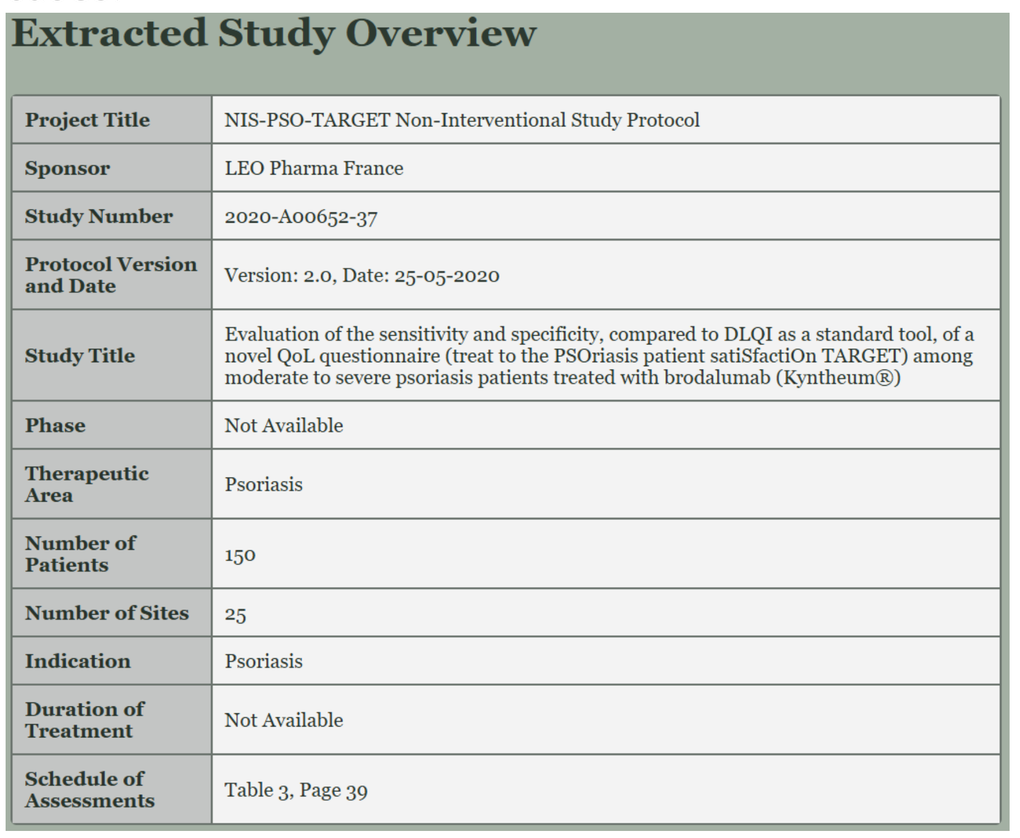


Issues Resolved

- Inaccurate outputs from GPT-4. We solved it by iteratively testing prompts on real clinical trial protocols.

- Difficulty matching text (GPT) with the tables (Adobe). Different formats for the data and different names. We solved it using a vector similarity score to give a “match” score between two entries. We also used preprocessing to normalise text before comparison.

- Generalising each protocol. Every protocol uses different layouts, terminology and formats. We had to add flexibility to the prompts and validation rules to handle all the variation. We also used threshold based matching and fallback rules to handle edge cases.



Technologies

- TimelineXtract is a Django application running on Render with a react front end running on Vercel.

- Data extracted with GPT4 and Adobe.

- Users, PDFs, Queries, Tables and Outputs are all stored in MongoDB

Ollscoil Chathair Bhaile Átha Cliath
Dublin City University

mongoDB

Timeline Xtract