

CASE STUDY

InformaticTech Creates an Integrated Solution for Endpoint Evaluation of Protocols and Clinical Study Reports



Client Overview

The client needed an efficient and flexible endpoint evaluation method for protocols and clinical study reports (CSRs).

The Challenges

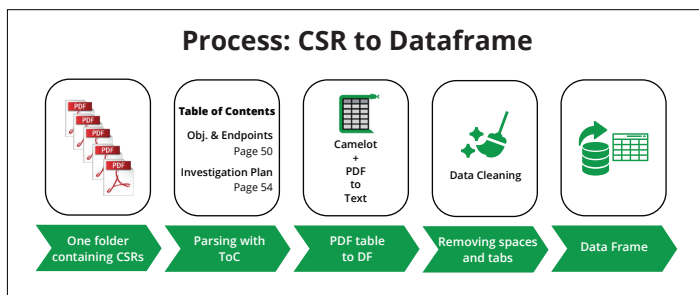
The main challenge faced by the client is determining which protocol endpoints consistently provide value. Additional endpoints bog down the evaluation process, resulting in:

- Added site burden
- Added organizational burden
- Increased R&D spending
- Longer timelines

Due to the length of documents and file types, it was difficult to create a feedback loop from the final clinical study report back to protocol.

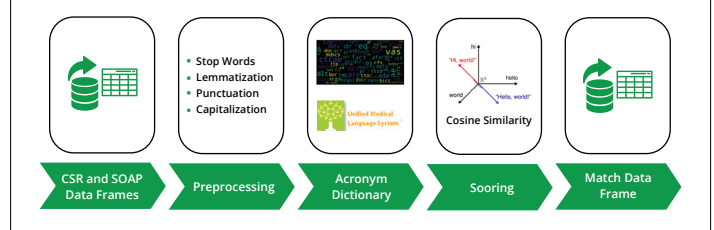
The Solution

Informatic employed two processes in the overall evaluation. First is the conversion of the CSR into a dataframe, and second is the comparison between CSR and Protocol.



Informatic tried out different natural language processing options and methodologies, eventually landing on a robust and scalable solution. This involved the use of the Camelot Python library to reliably extract tables from CSR PDFs. Camelot's two-step process allows the tool to reliably digest tables within pdfs. First, the tool identifies the table boundaries, and then it extracts text from the cells.

Process: CSR to Protocol Comparison






The comparison between CSR and Protocol begins with the preprocessing of the data frames and the use of an Acronym Dictionary. The documents are then scored using a Cosine Similarity metric to determine how similar they are. Informatic determined that the scoring system allows for more flexibility, including subject matter expert overrides, in the final tool.

Among the key takeaways from the results are:

- The scoring methodology shows all endpoints that have any level of match.
- It is possible to build automated prompts for fast and easy human review
- This method gives users the ability to determine project-specific thresholds.

Outcomes Delivered

-  **Optimization for specific use cases**
-  **Scalability and expandability**
-  **Compatibility with more "advanced" machine learning models**