Mining Experimental Conditions for HPLC-MS/MS Assay for Pharmacokinetics using NLP

The Data Mine and Merck Corporate Partnership

Figure 1: Introduction

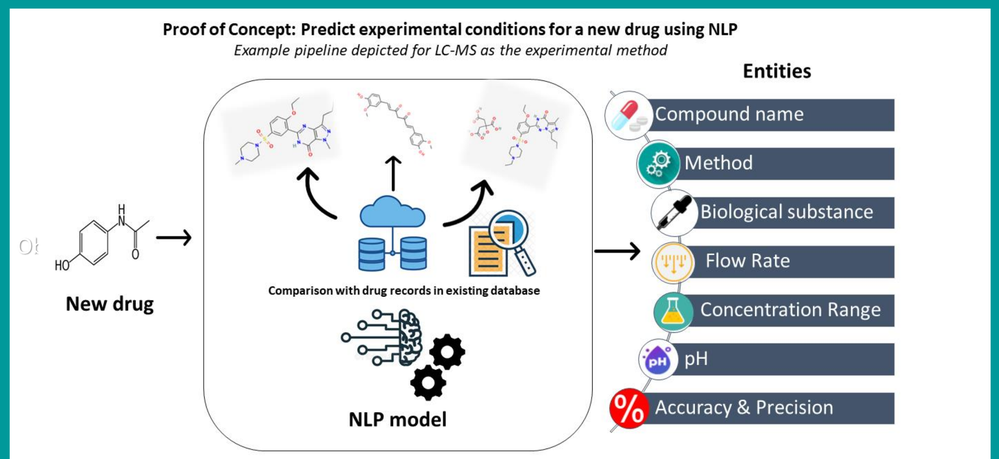
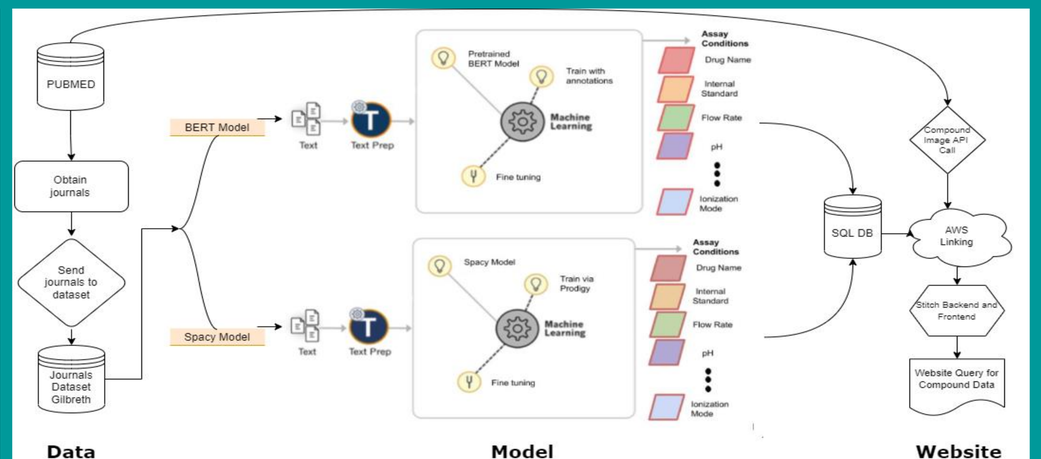


Figure 2: Product Workflow



Future Work

- Building a pipeline to extract full-texts of articles
- Connecting to Merck's internal databases
- Similarity search for compound structures
- Creating a quantitative metric to evaluate the quality of article

Acknowledgements

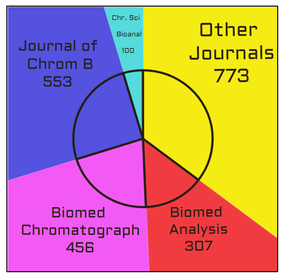
Thank you to Merck for their continued support of the The Data Mine. We extend the greatest amount of appreciation to our mentors Dr. Xiang Yu, Dr. Rajesh Desai and Dr. Gregory Bryman. Through their support, our project continues to grow and bring success.

Thank you to The Data Mine and the accompanying staff, specifically Dr. Mark Ward, Maggie Betz, Kevin Amstutz, Norma Grubb and Shuennhau Chang.

Data Acquisition

We developed a pipeline for obtaining article abstracts and full texts from journals listed in the PubMed:

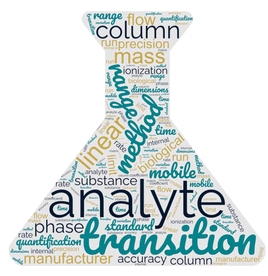
Fig 3: Source Journals Distribution



Using MESH indexing and the PubMed API allowed us to mine from the highest quality journals.

64% of the papers come from high impact Bioanalysis papers

Fig 4: Word Cloud of Experimental conditions



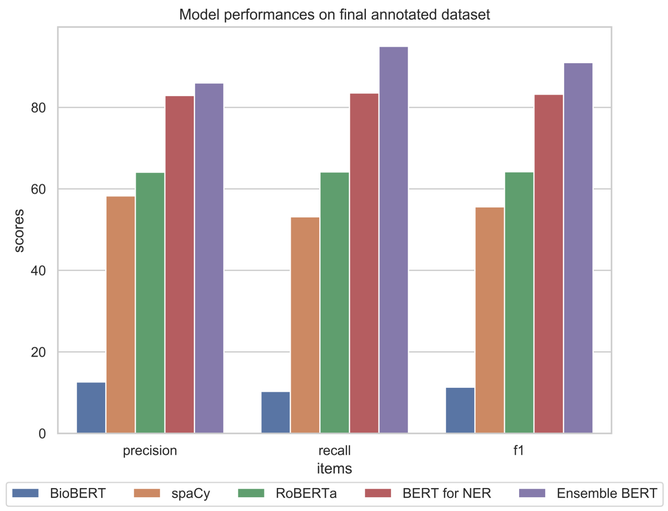
Automated framework to efficiently annotate 5 out of the 20 experimental conditions makes a day's worth of work a matter of seconds.

64% of papers have 7 or more complex experimental conditions

Model Development

Trained, evaluated and compared multiple NLP models to extract assays from abstracts with high accuracy

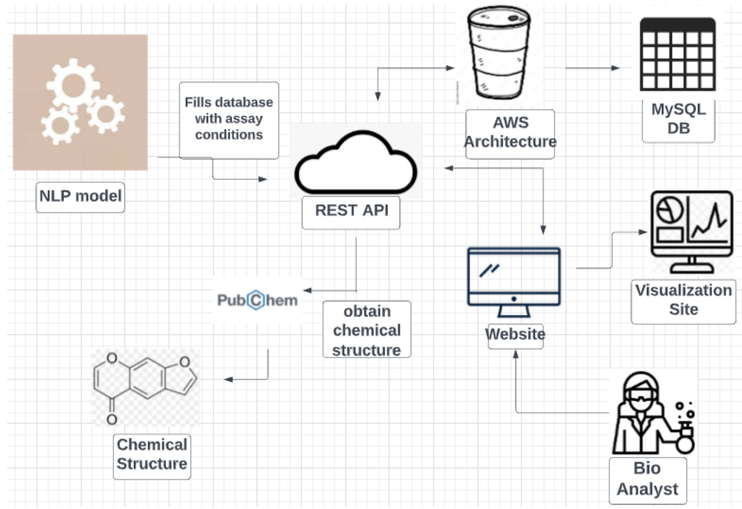
70% improvement in accuracy from existing pre-trained BioBERT model



Web Development

Easy to use website built for Bio-Analysts to query the curated assays to find their information quickly and efficiently

Fig 5: Software Architecture for Website



Miguel Alarcon – Vaishakh Deshpande - Kris Gurung - Milen Manoj - Shreya Puranik - Siddharth Srinivasan – Abhishek Shrinivasan – Charles Wang - Wei Xu - Qian Zhang
Dr. Gregory Bryman – Dr. Terri Bui - Dr. Rajesh Desai - Dr. Xiang Yu