Merck Lab Logistics – MSDS

Brian Haoqin Lu, Dennis Pham, Ravleen Kaur Chhabra, Shreyas Vaid, Sabharinath Saravanan

The **Crucible** software suite is a cutting-edge solution that processes and stores thousands of **material safety data sheets**, empowering scientists at **Merck** to accelerate the generation of critical **precautionary and hazards exposure** insights, and swiftly comply with **safety regulatory requirements**.

Problem

Scientists at Merck spend hundreds of hours manually populating and maintaining a list of thousands of safety data sheets (SDS) to create cover sheets for precautions and hazards exposure insights in order to comply with safety regulations.

Goals & Motivation

Our goal was to develop a fully automated system to populate and maintain a database of SDS documents, allowing scientists at Merck to programmatically generate cover sheets for their chemicals and potentially saving hundreds of thousands of dollars on wasted time annually.

Parsing Pipeline

Tungsten's SDS parsing pipeline is a flexible multi-stage system designed to accommodate multiple parsing paradigms.

Parsing Accuracy

To ensure that our SDS parser produces accurate output, we devised a way to measure how accurately a document has been parsed in accordance with the UN Globally Harmonized System. On average, our parser is 90% accurate for all sections. However, considering only the fields necessary to generate the cover sheet boosts accuracy to 100%, ensuring that no issues will occur for cover sheet generation.

User Interface Design

The original user interface (UI) was designed using Figma, the industry standard collaborative interface design tool. With Figma, we were able to create an interactive demo of the UI, so that we could receive feedback from potential end-users and reiterate on our design before developing the application.
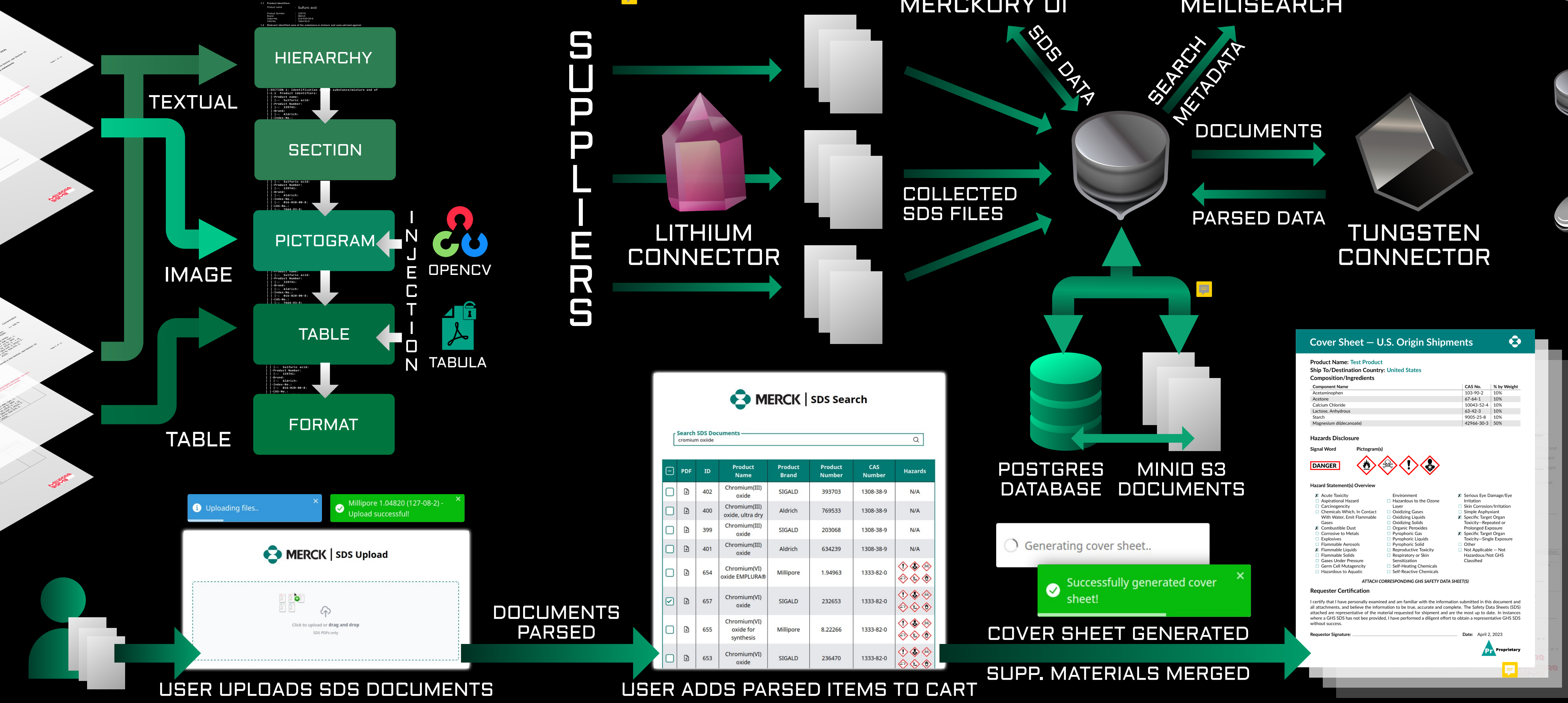
LITHIUM
SDS SCRAPER

An automation to scrape and download thousands of SDS from Sigma Aldrich.

TUNGSTEN
SDS PARSING SUITE

An open-source Python library built to parse and extract data from SDS.

SILICON
SDS DATA ACCESS

The backend API of the Crucible suite, providing programmatic access to SDS data.

MERCKURY
CRUCIBLE USER INTERFACE

The frontend interface built to allow end-users to search for chemicals and generate cover sheets.

Conclusion

Throughout this past year, our team has made significant strides in the development of Crucible. Our software suite streamlines the process from SDS document gathering to cover sheet generation, eliminating the potential for human error and saving valuable time for Merck scientists, allowing them to focus on other important tasks. Moving forward, we hope to expand support to multiple SDS vendors and enhance the observability of our platform.

Acknowledgements

Thank you to Merck and The Data Mine for their support to our team. We would like to especially thank Terri Bui and Stephen Ma for their guidance throughout this past year as they led us through development.