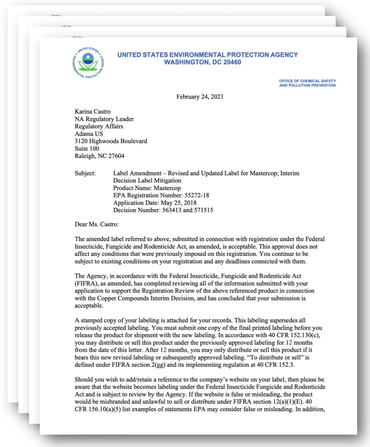
435- Project PESTICIDES: Data

sprout *Structured Pesticide Repository of Unified Tracking*

DEPARTMENT OF COMPUTER SCIENCE

Taehyeok Kwon, Terence Lo, Luke O'Neill, Sungbin Oh, David Perez, Cedric Yeo, Mathew Zheng

Problem Definition



UNITED STATES ENVIRONMENTAL PROTECTION AGENCY
WASHINGTON, DC 20460

OFFICE OF CHEMICAL SAFETY AND POLLUTION PREVENTION

February 24, 2021

Karina Castro
NA Regulatory Leader
Regulatory Affairs
Adama US
3120 Highwoods Boulevard
Suite 100
Raleigh, NC 27604

Subject: Label Amendment – Revised and Updated Label for Mastercop; Interim Decision Label Mitigation
Product Name: Mastercop
EPA Registration Number: 35272-18
Application Date: May 25, 2018
Decision Number: 563413 and 571515

Dear Ms. Castro:

The amended label referred to above, submitted in connection with registration under the Federal Insecticide, Fungicide and Rodenticide Act, as amended, is acceptable. This approval does not affect any conditions that were previously imposed on this registration. You continue to be subject to existing conditions on your registration and any deadlines connected with them.

The Agency, in accordance with the Federal Insecticide, Fungicide and Rodenticide Act (FIFRA), as amended, has completed reviewing all of the information submitted with your application to support the Registration Review of the above referenced product in connection with the Copper Compounds Interim Decision, and has concluded that your submission is acceptable.

A stamped copy of your labeling is attached for your records. This labeling supersedes all previously accepted labeling. You must submit one copy of the final printed labeling before you release the product for shipment with the new labeling. In accordance with 40 CFR 152.130(c), you may distribute or sell this product under the previously approved labeling for 12 months from the date of this letter. After 12 months, you may only distribute or sell this product if it bears this new revised labeling or subsequently approved labeling. "To distribute or sell" is defined under FIFRA section 2(gg) and its implementing regulation at 40 CFR 152.3.

Should you wish to add/retain a reference to the company's website on your label, then please be aware that the website becomes labeling under the Federal Insecticide Fungicide and Rodenticide Act and is subject to review by the Agency. If the website is false or misleading, the product would be misbranded and unlawful to sell or distribute under FIFRA section 12(a)(1)(E). 40 CFR 156.10(a)(5) list examples of statements EPA may consider false or misleading. In addition,

EPA

The U.S. EPA contains over **360K** pesticide-label documents—each packed with critical application and safety instructions for growers—but they exist only as inconsistent PDFs that are cumbersome to search, forcing farmers to comb through lengthy documents for dosage, timing, or safety guidance.

Our solution harnesses **state-of-the-art artificial intelligence** to automatically parse and extract important data—crop, target pest or disease, application rate, safety information … —and loads them into an **open-source database**. This AI-driven pipeline turns text into actionable insights, so growers can focus on farming rather than paperwork.

Methodology

Label Filtering → **PDF Parsing** → **AI Inference** → **Database Insertion**

STEP 1: Label Filtering

Filter down the EPA database entries into a subset of 5K pesticide labels. We narrow down to the scope defined by our client.

STEP 2: PDF Parsing

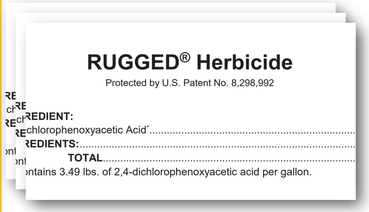
Transform the 5K PDF pesticide labels, including tables, into text files using PDF Parsers and OCR.

STEP 3: AI Inference

Take in the text files and extract out the pesticide information by inferencing a LLM (reasoning model).

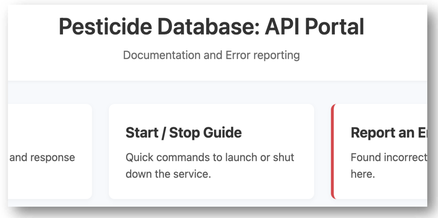
STEP 4: Database

Parse the JSON and upload the data into the database, where it can be accessed via a public API.



RUGGED® Herbicide
Protected by U.S. Patent No. 8,298,992
REDIENT:
chlorophenoxyacetic Acid*.....
REDIENTS:.....
TOTAL.....
ntains 3.49 lbs. of 2,4-dichlorophenoxyacetic acid per gallon.

PyMuPDF



Pesticide Database: API Portal
Documentation and Error reporting
and response
Start / Stop Guide
Quick commands to launch or shut down the service.
Report an E
Found incorrect here.

Our Goals

Empower farmers, researchers, and agricultural professionals with reliable insights on pesticide application, safety guidelines, and compliance.

90% Accuracy on automated extraction pipeline

AI Model Research

At first, we considered a variety of possible models to use, including models from OpenAI, Google, Anthropic, and others. Additionally, we experimented with open-source models from Hugging Face.

AI models aren't perfect, and many of the models we tested were subject to hallucinations and accuracy struggles, including missing or incorrect information. In the pursuit of an accurate and cost-effective solution, we developed a variety of parsing techniques, tested over a dozen models, and explored several prompt engineering strategies.

After many weeks of experimentation, we found OpenAI's newest reasoning model o4-mini was the most powerful in our price range.

OpenAI

Cost per label on average: 5.4 cents
Overall accuracy: >90% (based on internal experiments)

o4-mini