



Finds apps that shut down while people still wanted them

A plugin scraper pipeline that reads replacement requests and ranks the gaps worth building into.



Sytze P.
Solution Engineer

The problem

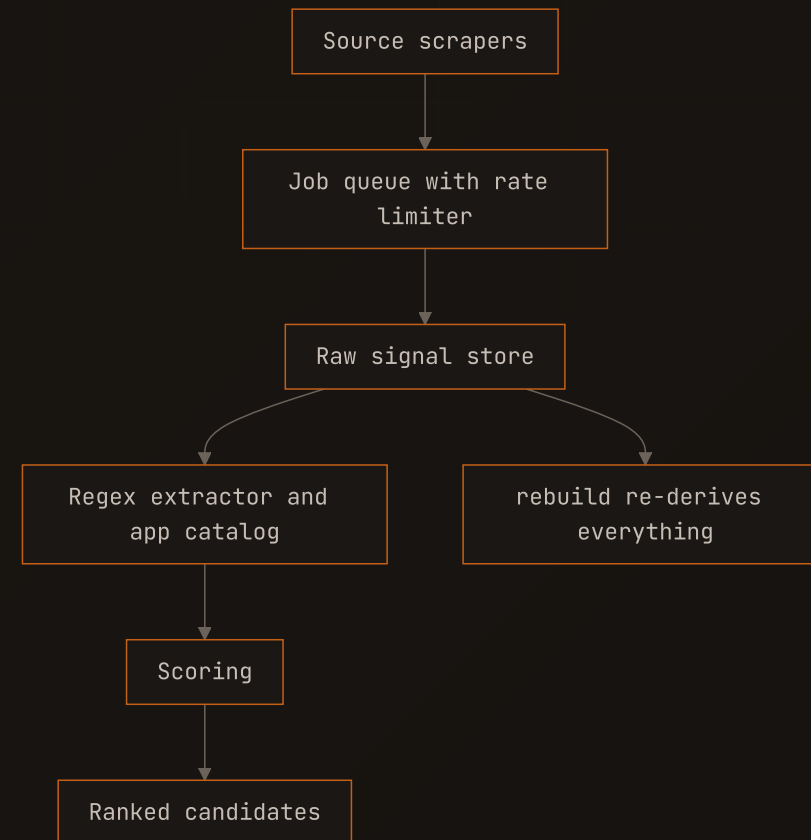
When a popular app shuts down, the people who relied on it go looking for a replacement in public. That demand is visible for a while, then it disperses and the opportunity closes with it.

Spotting it by hand means reading the same forums every day and remembering what you already read. Nobody sustains that for long.

Scrape to ranked candidate

Each source is an isolated scraper file dropped into a registry, so adding one never means refactoring the others. A job queue with rate limiting feeds them, and the store is SQLite in write-ahead mode with durable cursors, so a hard crash resumes instead of starting over.

Extraction is deliberately regex and a curated app catalog rather than a language model. These signals are short and formulaic, so a model would add cost and non-determinism without buying accuracy. A rebuild command re-derives every signal from the raw captures, which is the safety net for the day an extractor turns out to be wrong.



- HIGHLIGHTS

How it works

- SCRAPERS ARE DROP-IN

A new source is one file added to the scrapers directory. Nothing else in the pipeline changes to accept it.

- SURVIVES A HARD CRASH

Write-ahead logging and durable cursors mean an interrupted run picks up where it stopped rather than re-scraping from the start.

- WRONG EXTRACTOR IS NOT A LOST DATASET

Raw captures are kept, so a rebuild truncates the derived tables and replays everything through the fixed extractor.

Results & impact

no model in extraction

regex and a curated catalog instead

resumes after a crash

write-ahead store with durable cursors

fully re-derivable

rebuild replays every raw capture

• STACK

Tech stack

Python

pipeline and CLI

Playwright

browser-based collection

SQLite

durable store, write-ahead mode

Typer

command line interface

rapidfuzz

catalog matching

pytest

unit and end-to-end tests



Sytze P.

Solution Engineer

A plugin scraper pipeline that reads replacement requests and ranks
the gaps worth building into.

