



Datamam

Building a World Inspired by Data

A Case Study:

Custom Data Scraping Solution for Legal Research

October 2021

Overview

A major player in the field of legal research and analytics, engaged us to streamline their process of docket data collection from **more than 70 county websites** for their Meta-analysis project.

Their aim was to create a comprehensive, searchable database of all available cases from a specific list of county websites, with distinguished outcomes for analysis by their research team.

By supplying a steady, automated flow of structured data, we enabled the client's data team to streamline their tasks, optimizing their database population processes.

This meticulous strategy ensured the client was furnished with a precise, and readily analyzable database.



Challenges

During the project, our team faced several challenges:

- Structural Diversity: Diverse data structures across websites and timeframes
- Anti-Scraping Hurdles: Barriers like captchas and IP blocking
- Volume and Variation: Diverse standard formats and massive data volume
- Historical Complexity: Requirement of historical data with changing structure
- Client-Centric Formatting: Client-specific format requirements

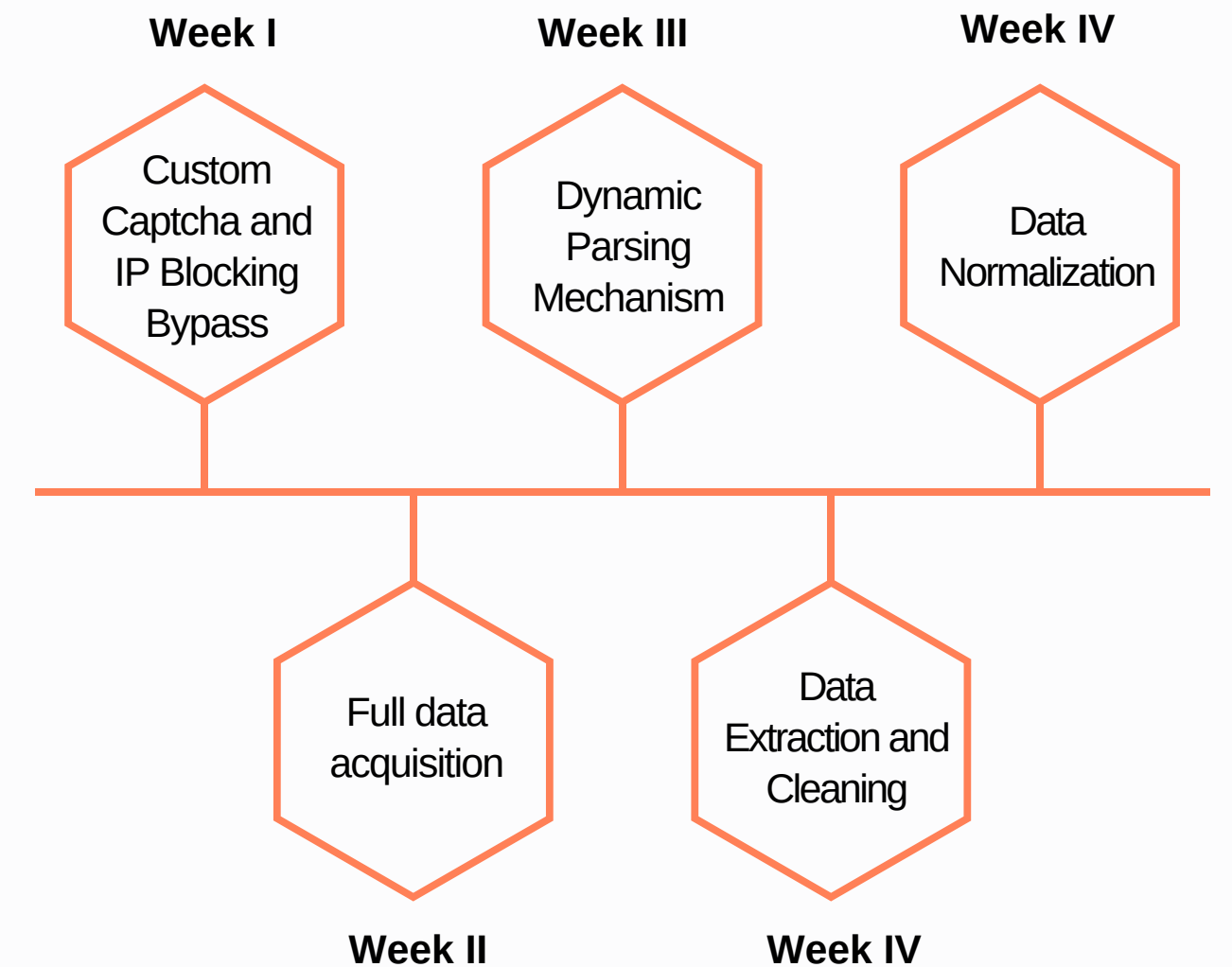


Solution

Datamam's custom solution addressed these challenges through:

- Custom Captcha and IP Blocking Bypass: Bypass techniques for captchas and IP blocks
- Full data acquisition: Search query tailoring to respect website limits
- Dynamic Parsing Mechanism: Dynamic parsing to accommodate varying docket structures
- Data Extraction and Cleaning: A robust data extraction system with meticulous cleaning
- Data Normalization: A process to normalize the data into the client's preferred format

Timeline



Impact

Operational Efficiency Boosted: Our solution transformed manual tasks, letting the research team dive into advanced analytics, and enhancing productivity. This also reduced potential errors, ensuring more accurate and time-efficient research processes.

Redefining Legal Research: Offering a structured database enabled deeper insights, promoting innovation and strengthening their industry position. This unprecedented access to organized data has set new benchmarks for quality in legal research.



Key Takeaways

The case underscores the importance of:

- **Personalized Approaches:** Tailoring solutions to specific problems ensures not only that the issue at hand is addressed but also that the solution seamlessly integrates with existing systems and processes. Custom solutions pave the way for more efficient, effective outcomes by catering directly to the unique challenges presented.
- **Consistency in Data Utilization:** Standardized data doesn't just mean a cleaner dataset. It means a dataset that is more easily manipulated, analyzed, and interpreted. By creating uniformity across diverse sources, we ensure that data is both accessible and actionable, facilitating better decision-making processes.
- **Automation Unlocks Potential:** While manual efforts might be stymied by vast data sets or inherent platform limitations, automation transcends these boundaries. Automated techniques are designed to handle large-scale data extraction, organization, and analysis, ensuring comprehensive coverage and accuracy, even when faced with constraints.

Contact



+1 650 405 0504



info@datamam.com



datamam.com



Menlo Park, CA 94025, USA

Thank You For Your Attention

