AUTOMATING DATA ACQUISITION WITH WEB CRAWLERS

Trexin helped a leading CPG company deploy Web scraping technology to automate a costly manual activity.

BUSINESS DRIVER

As part of a data warehouse project to enable our Client, a leading consumer packaged goods company, to create a fact-based selling case through analytics to drive retailer collaboration and profitably grow revenue, it was recognized that data acquisition was a costly and inefficient process. Historically, category management analysts spent up to 80% of their time each week manually acquiring point of sale (POS) data from retailers' Web portals. With some retailers preparing to provide daily POS updates, this necessary but time-consuming activity would only increase, further diverting analysts' attention from higher-value activities that could help the company better understand their sales and customers. The company's VP of IT challenged Trexin to find a way to automate this task.

APPROACH

Trexin first considered whether any of the retailers accommodated direct system-to-system integration via an application programming interface, which would be the ideal solution. Learning that only Web-based user interfaces were available, Trexin immediately shifted its focus to Web scraping technologies that could crawl through the set of retailer portals and systemically extract and process data in a fully automated manner. Trexin evaluated several vendors with advanced Web crawler technologies for their fitness against our Client's business requirements, overall cost of ownership, ease of maintenance, and experience scraping retail data.

After selecting the best vendor, Trexin built 20 Web crawlers that mimicked the data acquisition and loading function that was heretofore performed by four analysts. The crawlers, which require no human intervention, were configured to run on schedules to retrieve data related to stores, products, and sales, extracting data from the retailers' POS portals and delivering it to our Client's SFTP site where it could then be accessed by the data warehouse's conventional extract-transform-load (ETL) tools.

To build these Web crawlers, Trexin had to have a deep understanding of analysts' processes as well as the data they were using. Not only were we able to automate their data acquisition process, but we were also able to make changes to the data they were acquiring based on our industry expertise that would allow them to pull a cleaner data set that was more reflective of how competitors and retailers themselves look at POS data.

RESULTS

Automating the data acquisition process through web crawlers shifted analysts' focus from manual data pulls to value-add analyses. In addition to freeing analysts' time, the Web crawlers also allowed for a standardized process and view of the POS data, which allows for comparative analysis that were either previously impossible or too cumbersome to perform.

Web Crawlers Automate Daily Point of Sale (POS) Data Acquisition

A Costly and Inefficient Process

While working with our Client, a leading consumer packaged goods company, Trexin recognized a time-consuming gap tied to a weekly manual POS data acquisition which consumed 80% of the analysts' time. Trexin was tasked at finding a way to automate this process.

Approx. weekly time spent on manual POS data acquisition — 80%
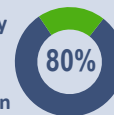
Crawling the Web for Data

Trexin selected a vendor experienced in Web crawling and who met our Client's business requirements

20 Web crawlers were built

Mimicked functions of **4** analysts

Web crawlers retrieved store, product and sales data from retailers' POS portals

The Web crawlers not only freed the analysts' time, they also allowed for:
- Standardized processes
- Easily viewed POS data
- Comparative analyses

CONTACT US

Products & Distribution Practice
pd@trexin.com
www.trexin.com

TREXIN
TRUST • EXPERIENCE • INNOVATION