algoscale

Building AI for a startup providing content optimization

Search engines favor pages that provide thorough overview of a topic that includes all fundamental subtopics, answer questions, and move user closer to satisfying their initial query. Nearly 75% users never go past the first page of a search engine. Therefore, SEO is pivotal in determining the rank of websites. The client wanted to ease the process of data extraction from their clients websites in order to improve their search performance. Algoscale leveraged Topic modeling, a complex form of AI to provide a solution that standardized their process and helped optimize their time.

The Client

Headquartered in Boston, U.S, the firm was founded in 2013. The client is a content marketing firm that uses AI to accelerate content planning, creation and optimization. The client is a finalist for the 2018 Red Herring and Top 100 North America Award.**The Challenge**

Search algorithm are getting progressively smart. Search engines engage models that measure the topical horizon of a page and not just the keywords. The client wanted suggestion on all the best suitable and relevant topics that was to be covered in the content for improving their SEO ranking to draw the attention of more targeted customer segments.

Huge amount of data had to be cleaned, organized and duplications were to be removed which led to increased complexity. Also the data was to be extracted from diversified sources and open – ended domains.

The Solution

Algoscale team suggested to build a web crawler that would standardize the process of data extraction from varied sources. Earlier the client needed custom crawlers for each new website which was inefficient. The basic idea was to standardize a crawler that could extract the data from different URLs and further the data would be cleaned, organized and stored using SOLR. Algoscale used Scrapy, a web-crawling framework, which was independent of the format of website and content, happened to provide more accurate solution resulting in saving of time for QA (Quality Assurance) analysis. The data extracted was junk free and quality test gave 90% accuracy.

Topic modeling, an established technique used to extract valuable topics from a corpus of data was used to determine specific keywords according to the needs, suggesting all the suitable and relevant topics that was to be covered in the content. MySQL database was setup to track the status of crawler i.e., it gets updated automatically on real time with the progress of crawler for the website and as well as for avoiding any type of repetition. Thereby, improving the SEO ranking by optimizing content around user intent to draw the attention of more targeted clients.

Benefits

Algoscale's solution helped client in meeting the deadlines because of the standardization of crawlers. Further, we increased the accuracy by 10-20% in extracting targeted topics thereby guiding the client to create world-class content which in turn drove more traffic to the website and earned links back to on-site content.

Technology Stack

Python, JAVA, Scala, Scrapy, MySQL & SOLR.