

## Scrape Airfare Data from Airline and Online Travel Sites

**Client:**A popular airline company from India

**Solution:**Site-specific crawl and extraction

**Domain:**Travel

**Challenge:**The client wanted to extract airfare data from competitor sites and other Online Travel Agency websites in order to power their competitor monitoring and pricing intelligence activities. Although the data was available on the target sites, the client lacked the technical infrastructure and expertise to programmatically access this data to feed the pricing engine and perform further analyses. The data was to be extracted at a frequency of 4 times a day and delivered in CSV format to their FTP server.

**The Solution:** After the client shared the specifics of their requirement such as source websites, URLs to be crawled and the data points to be extracted, we set up the crawlers for each site in the source list. The frequency was set as per the client's requirement and target data fields were departure time, flight duration, source and destination airport names and price. This requirement comes under our [site specific crawling](#) solution since all the sites in the source list followed different schema. We could complete the initial setup in 3 days and the first set of data containing about 100 k records was delivered promptly to the client.

### Benefits:

- The technically complex aspects of web crawling were completely owned and taken care of by our team
- It took only a few days for the crawler setup to be complete and the data flow was consistent thereafter
- We set up manual and automated monitoring systems for the target sites to detect changes
- Our extensive infrastructure could easily handle the large-scale data extraction at high frequency
- The client was able to use the data soon after it was delivered since it was clean and structured