



## Efficient Web Scraping Solution for R&D Institute

### Client Background

Our client is an R&D institute involved in the field of online platforms exploration, based in Germany. The aim of our client is to support businesses with relevant data and to provide the necessary scientific basis for the execution of business strategies.

### Business Challenge

Since most of the web-browsers don't offer the functionality to save a copy of viewed data, the only way to harvest the information is to manually copy and paste it. This makes a data extracting process a very tedious job which can take many hours or even days to complete.

In order to improve their data collection procedures, the client required an advanced solutions capable of scraping the information from various online platforms. By implementing new technologies into their operations, the client

wanted to gain more effective tools for having information collected in more efficient and faster manner. The implemented solution was expected to enhance customer's research approaches and to enlarge their market share.

## Project Description

The Db Devs team was provided with the list of over 100,000 user IDs on Wikipedia as an input data. Some of the users from the list already had their Wikipedia pages filled with data about their names, cities of origin, and native languages, but most of the users had their accounts blank. Another challenge was that some people instead of using free-format text while filling in their profiles, used wiki-syntax – a special software for page formatting, what made a web scraping of the provided data a particularly difficult matter.

The ultimate goal of the project was to match the users from the initially provided list with the list of over 400 German districts in order to have the relevant data and consider the project accomplished.

The engineers from Db Devs developed an efficient desktop web scraping application based on .NET codes with a built-in MS SQL database, that could automatically load and extract data from multiple pages of websites. With the click of a button the client could easily save the data available at website to a .csv file and store it to the home computer.

## Value Delivered

The engineers from Db Devs team managed to build a professional web data extraction software with a friendly, experience-driven interface. The provided solution was designed to perform the most labor-intensive operations automatically that helped the client greatly increase productivity and effectiveness of the web data scraping process.

## About SSA Group

SSA Group is a software development company that designs, implements, and supports cutting-edge digital solutions, enabling customers to unlock their business potential.

With over 10 years of experience, SSA Group creates innovative products and modernize complex legacy systems that shape today's digital and business landscape.

SSA Group is a trusted partner for the world's industry leaders, consistently turning their ideas into reality.

Contact us

Email: [contact@ssa.group](mailto:contact@ssa.group)

USA Toll Free: +1 866 263 99 03

Visit our website

<https://ssa.group/>



Reliable IT Partner  
for Your Business

GROUP