

Extensive web scraping experience



Product pricing and details



Finance



Social media



Business contacts



Media news



And many more

You name the source and data to extract and we will come up with a tailor-made solution adjusted to your needs!

We offer premium services for business

DevsData LLC is a small software & recruitment agency, with Google-level engineers and a vast network of senior expert contractors. Our priority is premium quality and business understanding.

WE WORKED FOR



CLIENT ABOUT US

"DevsData demonstrated a strong degree of proactivity, taking time to thoroughly understand the problem, business perspective and continuously suggesting performance and usability enhancements. I've worked with them on numerous projects over the last 3 years and I'm very happy."



Jonas Lee, Partner & Executive Vice President of Verus Financial LLC

MEDIA MENTIONS



Technologies:

- Python
- Requests
- MongoDB



Scraping data stored deep inside HTML

We worked on extracting data from **Filmweb - the second biggest movies database in the world**. It required as few requests as possible to collect all data about every movie/tv series on the website. The data was stored deep inside HTML. BeautifulSoup was used to collect essential information and parts of the website.

Technologies:

- Cron
- Python
- Requests
- BeautifulSoup
- Marshmallow
- MongoDB



Creating and maintenance of several small scrapers on short notice

We designed and maintained several small scrapers for a company project on short notice. Our task was to extract the data as quickly as possible and filter it to obtain only the essential information. **One of the projects was a Natural Language Processing scraping engine for a London-based hedge fund** - it scraped and scored news articles based on precise criteria given by the client.

Technologies:

- Python
- scrapy
- Google Cloud Platform (machines)



Scraping unstructured data from Wikipedia

We created a scraper running on Wikipedia to collect a data set regarding movies/television series and their cast. **The biggest threat in this project was that the website was non-structured, so links to other subpages could have been located everywhere.** Scrapy, which memorizes visited subpages and schedules pages to visit, was the most efficient technology to use.

Technologies:

- Cron
- Python
- Scrapy
- Marshmallow
- MongoDB



Boosting the efficiency of an existing scraping engine

We took part in the maintenance and modification process of the company's scraping engine. It was responsible for collecting profile data about people and companies from about ten confidential sources. The data had been purchased before, **so our task was to collect what was either not yet available to buy or to extend the possessed data.**

Technologies:

- Python
- scraPy
- Django
- PostgreSQL
- Scrapinghub platform (machines)
- Google Cloud platform (machines)



Gathering data from numerous websites

Our client needed to collect data on clothing products, with the main focus being their categorization and prices. There were about 30 websites with varying depth of information and protection against scraping.

Technologies:

- scraPy
- SQLite
- AWS (machine

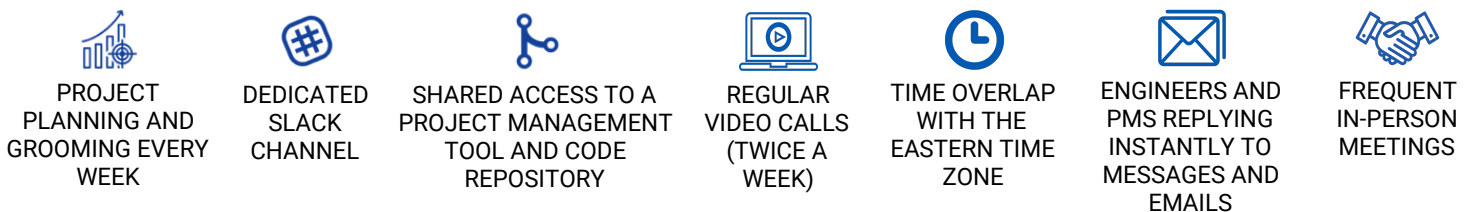


How we work

Most specialists usually **work remotely** from our European office, however, we are open for permanent relocation of selected engineers to NY or other US cities. For longer projects, we usually start full-time engagement with 2 weeks **onboarding locally** in the client's office in the US.

Communication is the key

We always make sure to be on the same page with our clients as we strongly believe that **communication is the key to fruitful cooperation**. During each project we enforce the following communication measures:



Our way of working is very transparent. We plan everything in advance and all the planning is visible in our PM tool so that the client can access it anytime.

More opinions from clients delighted with our performance

Experienced in large-scale projects

"Tom runs a unique IT consulting firm, DevsData. I have the pleasure of working with him as business partners on a big-scale digital transformation project for Varner - a European fashion Group with 11 thousand employees worldwide - they have assembled the core Machine Learning and Software team for the operation. I can personally guarantee their dedication, ability to understand the business needs and track down those exceptional people (we're very picky) wherever they are located. "



Nikolai Fasting, CEO of [Syncano, Inc.](#); Executive VP of Varner digital transformation

Mr. Nikolai is a serial entrepreneur with over 20 years of business experience. He has raised \$7M for his recent NY-based IT venture - Syncano - that is on a mission to popularize server-less technologies at scale. Recently Mr. Fasting has been involved as an executive VP in a challenging digital transformation project for Varner Group managing the budget of \$9M.

Flexible and dynamic

"The pace was very fast, with tight project milestones and deadlines as well as quick iterations. The machine learning approach was well thought out and the advantages, as well as drawbacks, were well-explained. The project was a huge success."



Angela Sy, Head of Strategy at [SkyCatch, Inc.](#)

Ms. Sy is a former Head of Artificial Intelligence at Skycatch, a venture-backed (\$46m) commercial drone startup based in San Francisco, California. Skycatch's goal is to build flying supercomputers (a.k.a. drones) to turn the physical world into an indexed database of quantitative information.

Selected tech stack we are familiar with





We have the **skills, experience and resources** to build complex technology systems for the most demanding clients. And most importantly - we like to **make things happen**.

LET'S TALK!

USA:

Tom Potanski, Owner

tpotanski@devsdata.com

(917) 267 8830

Europe:

Martha Karkosik, Account Executive

martha@devsdata.com

+48 606 331 020