Case Study

Personalized News Aggregator

Building a user-friendly news aggregator platform from scratch.

PYTHON   MONGODB   DOCKER   DJANGO   aiohttp
JQUERY   THEANO   SPHINXSEARCH
Case Info

Location:
Netherlands

Industry:
Information services

Partnership period:
December 2016 - July 2017

Team size:
2 people

Team location:
Kharkiv, Ukraine

Services:
Big Data development, Machine Learning, Web development, Data Science, Cloud architecture, Python development

Expertise delivered:
Big Data solutions, training ML models, Rapid detection system for checking the news source updates

Technology stack:
Python, Django, Docker, jQuery, Theano, Aiohttp, MongoDB, SphinxSearch
Client’s goals

The customer needed a platform for automated collection, verification, deduplication and classification of industry-specific news from a variety of credible sources. These news feeds would be both shown on a web portal, and/or could be embedded into any website.

The system should have had the following features:

- The data set must be **frequently updated and kept relevant**
- The news must be **categorized** to simplify the navigation and personalize the user experience
- **Only the primary source of the news** must be listed, not the duplicates
Project requirements

The client wanted each customer to be able to form a **unique, personalized news feed**, based on their preferences and search filters. IT Svit had to provide the following features:

- **The news data must be relevant**, meaning the data set must be frequently updated
- **The news must be categorized** to simplify the navigation and personalize the user experience
- To be valuable, **only the primary source of the news** must be listed, not the duplicates
- The tools must have **low hardware requirements**

Project results

IT Svit has built a powerful, flexible and highly performant **news aggregation platform**. Our **subscribers could embed our news feed code into their websites** and have the latest niche news related to their industry shown on respective websites.
Challenges resolved

IT Svit team started by building a highly-performant web scraper tool using aiohttp library. The data gathered was stored to MongoDB. We created several Machine Learning models using Theano and selected the one that provided optimal performance. We used these Machine Learning algorithms to filter out the irrelevant and/or unwanted content.

The platform front-end was built using jQuery, the back-end instances were containerized and automatically scaled using Docker. We used SphinxSearch to allow the customers to create their personalized news feeds. As the product was intended to be used as a plugin on any website, we had to make it both simplistic and highly configurable.

- We created a huge base of news sources worldwide, split by industries, locations, spheres of interest
- This base was frequently and rapidly updated, using a highly performant tool that swiftly analyzed the article content to check if it was updated
- The platform was able to create a personalized news feed for any customer, based on their preferences and settings
- The aggregator could be easily embedded into any website as a plugin with simple configuration

"IT Svit was very professional, and they delivered exactly what was needed. The platform they developed seemed like reading my mind and was able to build a relevant news feed based on few search requests and preconfigured filters. It helped many website owners empower their portals with highly-relevant niche-specific news feeds.

Dennis M, CTO
at NewsBuzz"
For more than 12 years IT Svit helps the companies worldwide innovate and overcome their challenges. We have completed more than 600 projects, delivered successful projects for 6 Fortune 500 companies, helped 16 startups secure the investment and develop awesome MVPs. Our team of 50+ specialists provides services that earn 4.9 out of 5 customer satisfaction on Clutch.co.

IT Svit specializes in DevOps services, Big Data technology, Machine Learning, bespoke Blockchain platforms, full-cycle services for startups, web development and QA.