

INTISARI

PENGEMBANGAN APLIKASI *WEB SCRAPING* BERBASIS *CLOUD* UNTUK MYTENS GOBEYOND MENGGUNAKAN SELENIUM DAN INTEGRASI GOOGLE SPREADSHEET

Aura Nisa' Hidayat

21/482690/SV/19983

Kebutuhan akan tingginya volume data operasional sehingga membutuhkan proses yang cepat, akurat, dan efisien. PT Telkom Indonesia selama ini masih mengandalkan proses manual dalam pengumpulan data dari *platform* internal perusahaan yaitu MyTEnS GoBeyond. Untuk mengatasi permasalahan tersebut, proyek ini bertujuan untuk mengembangkan sistem otomatisasi pengambilan data berbasis teknik *web scraping*, yang diimplementasikan menggunakan bahasa pemrograman Python dengan *library* Selenium. Sistem ini memanfaatkan layanan *cloud computing* dari Amazon Web Services (AWS) dengan menjalankan sistem pada server virtual Amazon EC2 berbasis sistem operasi Ubuntu, penjadwalan sistem dengan cron job, serta menggunakan Amazon CloudWatch untuk monitoring server. Data hasil *scraping* kemudian dikirim secara otomatis ke Google *Spreadsheet* melalui integrasi Google Sheets API menggunakan autentikasi OAuth 2.0. Hasil dari penelitian ini menunjukkan bahwa sistem berhasil melakukan *scraping* secara otomatis setiap hari, dengan proses *login*, navigasi halaman, verifikasi data, hingga pengunggahan ke *spreadsheet* berjalan lancar dengan rata-rata waktu 119 detik. Sistem ini terbukti dapat meningkatkan efisiensi waktu pengambilan data >90% dibandingkan proses manual sebelumnya, meminimalisir resiko *human error*, serta mendukung pemantauan *real-time* melalui *spreadsheet*. Dengan demikian, sistem yang dikembangkan mampu menggantikan proses manual secara efektif dan memberikan nilai tambah dalam proses pengumpulan data operasional harian.

Kata kunci: *Web scraping*, Selenium, Amazon EC2, Komputasi Awan, Otomatisasi Data.

ABSTRACT

DEVELOPMENT OF CLOUD-BASED WEB SCRAPING APPLICATION FOR MYTENS GOBEYOND USING SELENIUM AND GOOGLE SPREADSHEET INTEGRATION

Aura Nisa' Hidayat

21/482690/SV/19983

The increasing volume of data in corporate operations demands a fast, automated, and error-minimized reporting system. PT Telkom Indonesia has been relying on a manual process in collecting data from the company's internal platform, MyTEnS GoBeyond. To overcome these problems, this project aims to develop a data retrieval automation system based on web scraping techniques, which is implemented using the Python programming language with the Selenium library. This system utilizes cloud computing services from Amazon Web Services (AWS) by running the system on an Amazon EC2 virtual server based on the Ubuntu operating system, scheduling the system with a cron job, and using Amazon CloudWatch for server monitoring. The scraped data is then sent automatically to Google Sheets through Google Sheets API integration using OAuth 2.0 authentication. The results of this research show that the system succeeds in scraping automatically every day, with the login process, page navigation, data verification, and uploading to the spreadsheet running smoothly with an average time of 119 detiks. This system is proven to be able to increase the efficiency of data retrieval time >90% compared to the previous manual process, minimize the risk of human error, and support real-time monitoring through spreadsheets. Thus, the developed system is able to effectively replace manual processes and provide added value in the daily operational data collection process.

Keywords: Web scraping, Selenium, Amazon EC2, Cloud Computing, Data Automation.