

Desarrollo de una Herramienta Web Scraping que Gestione Datos para la Industria automotriz

Ing. Emanuel Cruz Chávez¹, Dr. José Rafael Mendoza Vázquez², Dra. Irma Delia Cuevas Rojas³, Dr. Sergio Javier Torres Méndez⁴ y M.C. Vicente Ramírez Palacios⁵

Resumen— En este artículo, se presenta una aplicación programada en Python para la búsqueda específica de información en sitios web. La aplicación emplea las bibliotecas de Python, BeautifulSoup, Selenium y PyAutoGUI para buscar y presentar en forma automatizada información de interés, se complementan además con la biblioteca Pandas y el modulo os para un manejo más óptimo de la información local. Finalmente se aborda un ejemplo implementado en la gestión de datos de diseño de arneses eléctricos automotrices.

Palabras clave—Web Scraping, Python, Selenium, BeautifulSoup, Industria Automotriz.

Introducción

La administración de información de componentes automotrices a través de software de ingeniería es un proceso crítico de la industria, la cantidad de datos que se maneja es muy grande y se actualiza constantemente. Se identifican dos dificultades en este proceso: la primera, el alto costo de software de gestión de datos para el diseño automotriz y la segunda, el lento intercambio de información entre los usuarios y fabricantes quienes poseen la información actualizada. En la web existe una gran cantidad de información de componentes automotrices publicada por los fabricantes, por ello, una combinación de herramientas de Python, como BeautifulSoup, Selenium y PyAutoGUI permiten recopilar datos estructurados de estos sitios web. Así se genera información precisa y detallada, se ahorra tiempo y esfuerzo al actualizar las bases de datos. En este artículo se presenta el Web Scraping como técnica automatizada para la extracción de datos disponibles en sitios web empleando las bibliotecas mencionadas, se presenta un ejemplo del desarrollo de esta herramienta para el diseño de arneses eléctricos automotrices.

Descripción del Método

Para el desarrollo de la herramienta, se utilizó una metodología de cinco etapas. Las etapas contempladas son: la definición del problema, determinación de requerimientos, programación de la herramienta; pruebas de la herramienta y ajustes finales. La metodología descrita, se presenta en las siguientes secciones



Figura 1. Metodología para la construcción de una aplicación de gestión de datos para la industria automotriz.

¹ El Ing. Emanuel Cruz Chávez es estudiante de la Maestría en Ingeniería Electrónica en el Tecnológico Nacional de México /Instituto Tecnológico de Puebla im07220841.62@puebla.tecnm.mx.

² El Dr. José Rafael Mendoza Vázquez es Profesor del Departamento de Eléctrica y Electrónica y Profesor de la Maestría en Ingeniería en Electrónica en el Tecnológico Nacional de México/Instituto Tecnológico de Puebla rafael.mendoza@puebla.tecnm.mx

³ La Dra. Irma Delia Rojas Cuevas es profesora de la carrera de Ingeniería en Logística y de la Maestría e Ingeniería Electrónica en el Tecnológico Nacional de México Campus Puebla, México irma.rojas@puebla.tecnm.mx

⁴ El Dr. Sergio Javier Torres Méndez es Profesor del Departamento de Metal-Mecánica y Profesor de la Maestría en Ingeniería en Electrónica en el Tecnológico Nacional de México/Instituto Tecnológico de Puebla sergio.torres@puebla.tecnm.mx

⁵ El M.C. Vicente Ramírez Palacios es Profesor del Departamento de Ingeniería Eléctrica y Electrónica y Profesor de la Maestría en Ingeniería en Electrónica en el Tecnológico Nacional de México/Instituto Tecnológico de Puebla vicente.ramirez@itpuebla.edu.mx