

Un aporte al sistema universitario público argentino: utilización de Web Scraping para la facilitación de la oferta académica de cursos

Matías Yael Cortés; Thiago Daniel Perez; Patricia Valentina Zacarías Fuentes;

Paula Virginia Zacarías Fuentes

Docentes: Valeria Lasagna; Julieta Rivero; Romina Istvan

Laboratorio de Ingeniería en Sistemas de Información, LINES UTN FRLP

Av. 60 s/n° esquina 124, La Plata, Buenos Aires, Argentina

{mcortes; thiagodanielperez; patriciavalentinazacariasfuentes}@alu.frlp.utn.edu.ar ,

paulavzf2004@gmail.com

Área Relacionada

Trabajos estudiantiles

Resumen

El proyecto Automatización de la Oferta Académica mediante Web Scraping del Laboratorio de Ingeniería en Sistemas de Información (LINES) de la Facultad Regional La Plata se enfoca en la centralización de la oferta de cursos de extensión de universidades públicas argentinas. Esto se logra mediante la integración de dicha información en un sitio web unificado. El objetivo principal de esta plataforma es mejorar la visibilidad y accesibilidad de las actividades de extensión universitaria. Este trabajo detalla la experiencia obtenida por quienes participaron en el proyecto, destacando las instancias de aprendizaje colaborativo con estudiantes de cursos superiores, egresados y mentores. A lo largo del proceso, se logró la incorporación de nuevas herramientas tecnológicas y la aplicación de conocimientos preexistentes para la creación de un sitio web.

Palabras Claves

Web Scraping; Cursos; Universidad Pública Argentina; Python; HTML; Docker.

Introducción

La Universidad Tecnológica Nacional (UTN) cuenta con un gran abanico de sedes distribuidas a lo largo y ancho del país, cada una ofreciendo distintos cursos de extensión y capacitaciones. Sin embargo, esto implica que cada regional gestione sus ofertas académicas mediante su propio sitio web, generando así un problema de descentralización digital de la oferta académica. Este desafío no es exclusivo de la UTN: otras universidades públicas argentinas también enfrentan dificultades similares, con sistemas fragmentados que dificultan el acceso unificado a la información sobre cursos y actividades de formación continua.

Esta fragmentación digital dificulta el acceso fácil y ordenado a la información, especialmente para estudiantes, docentes o cualquier interesado en general. La falta de estandarización en los formatos de publicación genera una experiencia de búsqueda ineficiente, desordenada y, en muchos casos, frustrante para el usuario.

En este contexto, se comenzó a trabajar en una solución estratégica de automatización basada en el uso del Web Scraping como metodología de obtención de datos debido a la ausencia de APIs públicas para recopilar los datos que necesitamos. Utilizando el raspado de datos, se extrae y filtra la información publicada por distintas regionales, manteniendo únicamente los datos relevantes para mostrarlos de manera ordenada y centralizada en nuestra web.

Este desarrollo constituye una experiencia académica y formativa significativa. El proyecto involucra a estudiantes, egresados y docentes de Ingeniería en Sistemas de Información, y promueve un espacio de aprendizaje colaborativo, incorporación de nuevas herramientas tecnológicas y aplicación de conocimientos en un contexto real.

Documentar cada paso del proceso no solo facilita la incorporación de nuevos miembros al equipo, sino que también visibiliza el rol de la ingeniería como herramienta de mejora concreta para el sistema universitario público, a través de soluciones tecnológicas de impacto institucional.

Objetivos de la experiencia/investigación

El presente trabajo describe una experiencia de desarrollo tecnológico en un entorno institucional real, donde se integran saberes técnicos, aprendizajes colaborativos y valores sociales.

La iniciativa se centra en aplicar la técnica de web scraping para automatizar la recopilación y centralización de la oferta académica de cursos de extensión de universidades públicas argentinas, buscando no solo mejorar su accesibilidad y visibilidad, sino también diseñar soluciones tecnológicas inclusivas y escalables que contribuyan a democratizar el acceso a la educación pública.

Desarrollo

La oportunidad de aplicar los conocimientos de la cursada en un entorno que simula el mundo profesional, se llevó a cabo a través de la participación en el laboratorio LINES. Impulsado por una propuesta de las docentes Valeria Lasagna y Romina Istvan, de la cátedra Seminario Integrador perteneciente a la carrera Ingeniería en Sistemas de Información, nos sumamos a la iniciativa de centralización de la oferta académica con el objetivo de llevar los conocimientos ya adquiridos a la práctica.

El proceso de integración del grupo al proyecto fue gradual, comenzando con una serie de reuniones periódicas con el equipo del laboratorio. Estas primeras sesiones fueron fundamentales, ya que permitieron comprender mejor el estado del proyecto, así como interactuar con estudiantes más avanzados, egresados y mentores en un entorno de aprendizaje colaborativo. Durante esta etapa inicial se acordó, en conjunto, el alcance de la participación.

Se comenzó utilizando técnicas de web scraping, que consiste en la extracción, el rastreo y análisis de las páginas web para identificar y obtener contenido. En este caso se utilizó para acceder a la información dispersa en los distintos portales de las distintas facultades de la UTN. Al continuar el desarrollo, la idea fue mejorar y avanzar en el sistema, sin interrumpir su funcionamiento ni modificarlo por completo. Además, incorporamos tecnologías y recursos tales como:

1. **Desarrollo en Python para Web Scraping:** Más allá de la sintaxis del lenguaje, el desafío principal fue aplicar Python para construir *scrapers* funcionales y robustos. Esta tarea nos obligó a pensar de manera analítica, depurar código de forma eficiente y diseñar algoritmos capaces de adaptarse a diferentes fuentes de datos, consolidando así nuestra competencia como programadores.
2. **Control de versiones con GitHub:** Una organización rigurosa es clave en trabajos de desarrollo colaborativos. Por ello, la utilización de GitHub como sistema de control de versiones fue un pilar en nuestra experiencia. Esta práctica no solo garantizó la integridad del proyecto, sino que nos brindó una experiencia directa con los flujos de trabajo estándar en la industria del software.
3. **Contenerización con Docker:** Como parte de los contenidos avanzados de la cátedra, la implementación de Docker fue otro de los hitos en nuestro desarrollo. El uso de Docker nos permitió empaquetar la aplicación y sus dependencias, asegurando que el sistema se comportara de la misma manera en cualquier entorno y facilitando la futura inserción de nuevos miembros al equipo.

En paralelo al desarrollo técnico, nuestro rol también abarcó la investigación del marco teórico y la documentación formal del proyecto. Esta dualidad de responsabilidades nos proporcionó una visión integral,

donde no solo éramos ejecutores, sino también analistas y comunicadores, promoviendo así la experiencia formativa que buscábamos.

Conclusiones

La participación en el proyecto ha representado una experiencia formativa que fue más allá que el desarrollo técnico. En todo este trayecto, no solo alcanzamos los objetivos de aplicar los conocimientos que adquirimos, sino que hemos cumplido la capacidad que tiene un ingeniero en sistemas para poder proporcionar soluciones de alto impacto dentro de nuestro propio ecosistema universitario.

Esta experiencia educativa no solo termina en la entrega de una herramienta funcional para aportar valor al sistema universitario público, sino que también sienta bases para futuras expansiones y mejoras. La plataforma está desarrollada de manera que deja la puerta abierta a la incorporación de nuevas universidades, la implementación de funcionalidades avanzadas y optimización continua de los algoritmos de extracción de datos.

En conclusión, este proyecto ha sido un nexo fundamental entre la teoría académica y la práctica profesional, demostrando que, a través de la ingeniería y la colaboración, es posible construir herramientas que facilitan el acceso a uno de los pilares más importantes: la educación pública.

Referencias bibliográficas

1. Duy, L. (2024). *Web scraping of university rankings and data analysis using Python* [Trabajo de grado, Theseus]. https://www.theseus.fi/bitstream/handle/10024/851908/Duy_Le.pdf?sequence=2&isAllowed=y
2. International Journal of Advanced Science and Computer Applications. (2021). *Web scraping: State of the art, techniques, approaches and application*, 2(3), 11–19. <https://ijasca.zuj.edu.jo/PapersUploaded/2021.3.11.pdf>
3. Khder, M. A. (s.f.). *A review of web scraping technologies and their applications*. <https://repository.udistrital.edu.co/server/api/core/bitstreams/f1f64f24-7be7-4065-b5c1-f5117a50b1cf/content>
4. Kuder, M. A. (2021). *Web scraping: State of art, techniques, approaches and application*. *International Journal of Advanced Science and Computer Applications (IJASCA)*, 2(3), 11–19. <https://ijasca.zuj.edu.jo/PapersUploaded/2021.3.11.pdf>

5. Lawson, R. (2015). *Web scraping techniques: How to extract data from websites* [Tesis de maestría, Malmö University]. CORE. <https://core.ac.uk/download/pdf/237084705.pdf>