



UNIVERSIDAD DE GUADALAJARA

Centro Universitario de Los Valles

Secretaría Académica

Licenciatura en Tecnologías de la Información

Propuesta de Proyecto para Titulación

<u>Título del proyecto:</u> “Repositorio institucional para trayectoria y producción de docentes del CUValles utilizando técnicas de web scraping”	
<u>Tipo de proyecto:</u> Desarrollo tecnológico	<u>Financiamiento:</u> Ninguno
<u>Proyecto propuesto por:</u> Dr. Rodolfo Omar Domínguez García; Dr. Omar Alí Zatarain Durán; Dr. Miguel Ángel de la Torre Gómora	
<u>Número de estudiantes que pueden participar en el proyecto:</u> 1-2 estudiantes con capacidad autogestiva e interés por el aprendizaje continuo.	
<u>Descripción del proyecto:</u> <p>En el Centro Universitario de Los Valles se lleva de manera manual un registro de la trayectoria académica de los profesores adscritos a los distintos departamentos. Este proceso requiere de un esfuerzo constante por parte de profesores y administrativos, y comúnmente se tienen registros no actualizados y con errores. Por otro lado, existen plataformas como el curriculum de PRODEP y el curriculum vitae único de CONACyT, las cuales son alimentadas constantemente por cada uno de los profesores.</p> <p>En este proyecto se plantea el desarrollo de un sistema que lea los datos de trayectoria y producción de profesores, a partir del archivo PDF y/o word generado por los sistemas de PRODEP y CONACyT. Típicamente, a las técnicas que extraen datos de páginas web o documentos PDF se conocen como web scraping.</p> <p>Ligas de interés y material en línea:</p> <ul style="list-style-type: none">• https://es.wikipedia.org/wiki/Web_scraping• https://books.google.com.mx/books?id=TYtSDwAAQBAJ&lpg=PT8&ots=v0B_sFlkil&dq=web%20scraping%20python&lr&pg=PT8#v=onepage&q=web%20scraping%20python&f=false• https://link.springer.com/content/pdf/10.1007/978-1-4842-3582-9.pdf	
<u>Requerimientos de software y hardware:</u> Computadora, acceso a bases de datos (wdg.biblio.udg.mx), recursos metodológicos de revisión sistemática.	
<u>Resultados y productos esperados:</u> <ul style="list-style-type: none">○ Sistema funcional que extraiga información de los CVU de PRODEP y CONACyT, y almacene la información en una base de datos, para después generar reportes en web.○ Tesis en español o inglés que describa detalladamente el sistema propuesto y el proceso de desarrollo.○ [Opcional] Artículo de divulgación a presentar en conferencia nacional/internacional – (por ejemplo, CIMPS http://cimps.cimat.mx/ o LACCEI http://laccei.org/)	
<u>Tiempo estimado de duración:</u> De 6 a 12 meses de trabajo a tiempo parcial (entre 10 y 15 horas por semana por estudiante)	