

Análisis de la presencia de habilidades blandas en bolsas de trabajo usando web scraping en Python

Analysis of the presence of soft skills in job boards using web scraping in Python

José Eduardo Guevara Molina¹, María Azucena Flores Sánchez¹, Judith Pérez Marcial, Ana Claudia Zenteno Vázquez¹, María del Carmen Santiago Díaz¹, Gustavo Trinidad Rubín Linares¹, Yeiny Romero Hernández¹

¹ Facultad de Ciencias de la Computación, Benemérita Universidad Autónoma de Puebla, Avenida San Claudio, s/n. Puebla, México.
{jose.guevaramo, maría.floressanche}@alumno.buap.mx, {judith.perez, ana.zenteno, marycarmen.santiago, gustavo.rubin, yeiny.romero}@correo.buap.mx

Fecha de recepción: 20 de noviembre de 2023

Fecha de aceptación: 29 de abril de 2024

Resumen. Este trabajo se realizó con el propósito de determinar la demanda de habilidades blandas en publicaciones electrónicas de puestos de empleo del área de tecnologías de la información. Para la recolección de los datos se usaron técnicas de web scraping para obtener el contenido de las solicitudes y se usó el lenguaje de programación Python para su inserción a una base de datos. En total se obtuvieron 201 solicitudes de empleo de las cuales 155 fueron de www.jobatus.mx y 46 de www.occ.com.mx. Tras el análisis de dichas solicitudes y después de la limpieza de los datos, se llegó a la conclusión de que no existe un interés por parte de las empresas en colocar dichas habilidades blandas en sus publicaciones, sin embargo, estas se revisan en una fase posterior del proceso de contratación, a través de pruebas psicossomáticas o situaciones hipotéticas en entrevistas de trabajo u otros medios.

Palabras clave: Web scraping, Python, Soft Skills, Habilidades Blandas, Bolsa de trabajo.

Summary. This work was carried out with the purpose of determining the demand for soft skills in electronic publications of job positions in the area of information technologies. To collect the data, web scraping techniques were used to obtain the content of the requests and the Python programming language was used to insert them into a database. In total, 201 job applications were obtained, of which 155 were from www.jobatus.mx and 46 from www.occ.com.mx. After analyzing these requests and after cleaning the data, it was concluded that there is no interest on the part of companies in placing these soft skills in their publications, however, these are reviewed in a later phase. of the hiring process, through psychosomatic tests or hypothetical situations in job interviews or other means.

Keywords: Web scraping, Python, Soft Skills, Soft Skills, Job Board.

1 Introducción

La presente investigación tratará de dar respuesta a la siguiente pregunta: ¿Qué tan interesados están los empleadores en incluir habilidades blandas como requisito en sus bolsas de trabajo? La cuestión se generó por la necesidad de tener una mejor comprensión de las exigencias de los empleadores y quizá, analizar el requerimiento de fortalecer la fomentación de dichas habilidades ante un mercado laboral cada vez más competitivo.

Teniendo esto en mente, nuestro objetivo es identificar las habilidades blandas que solicitan los reclutadores al momento de publicar vacantes para empleos del área de las ingenierías asociadas a la computación en el estado de Puebla, en México. Para esto usaremos Web scraping, lo cual es una técnica para recolectar datos de una página web y su almacenamiento para posterior análisis [1], [2]. Por habilidad blanda se entiende como aquellas capacidades que mejoran el desempeño laboral, facilitar la movilidad interna, catapultar la carrera profesional y predecir el éxito laboral [3] por esta definición es que nuestra investigación es necesaria, debido que al conocer las habilidades más solicitadas, se pueden fortalecer las competencias que terminarán por mejorar la competitividad para puestos de trabajo de interés para el aplicante, además de no existir investigaciones centradas en la búsqueda de habilidades blandas en este campo.

2 Estado del arte del web scraping

Como se explicó antes, el Web scraping es una técnica para recolectar datos de una página web, para después ser almacenada para su estudio [1], [2].

Esta técnica de recolección de información es particularmente útil cuando se requiere analizar grandes volúmenes de información o en este caso, múltiples paginas donde se busca en los mismos fragmentos de código, por lo tanto, se han desarrollado algunas herramientas para llevarlo a cabo, entre estas están:

- Scrapy

Es un framework potente de web scraping para Python, donde los robots se definen como clases que heredan de la clase BaseSpider, que define un conjunto de 'urls de inicio' y una función de 'análisis' llamada en cada iteración web. Las páginas web se analizan automáticamente y los contenidos web se extraen utilizando expresiones XPath [4], [5].

- BeautifulSoup

Es una librería para extraer datos de documentos HTML y XML, proporcionando formas idiomáticas de navegar, buscar y modificar el árbol de análisis [6]

- Selenium

Es un driver web propiedad de selenium, el cual se usa para automatizar el navegador a través de comandos que se pueden acceder a través de la consola provista por el mismo [7].

Por la misma utilidad que tiene recolectar información ubicada en sitios web, múltiples trabajos han sido posibles gracias al web scraping. Citando un ejemplo, el trabajo de Christian Kempny y Patrick Brzoska [8] nos proporciona entornos en los que el web scraping puede ser aplicado mientras se investiga sobre los servicios de salud, así como si debería ser usado por todos los investigadores de la salud en sus trabajos.

Una aplicación más directa del web scraping se puede apreciar en el trabajo de Wang, Z., Zeng, J., Ran, H., Meng, W., Zhou, S., Davies, A. B., & Liu, C [9], donde se usó web scraping para monitorear la venta de especies de hormigas hacia el exterior de china, las cuales se podrían considerar especies invasoras y afectar el ecosistema, al mismo tiempo propone que el web scraping puede ser usado también para controlar el daño que estas especies pueden realizar verificando la procedencia y destino de las ventas.

Finalmente, como último trabajo citado en este estado del arte, citare la investigación de Macias, P., Stelmasiak, D., & Szafranek, K [10], donde se hacen predicciones inmediatas de la inflación de los precios de la comida desde 2009 en Polonia demostrando que las predicciones solas de los cambios de precio ya son efectivas para la predicción de futuras alzas.

Como nota adicional, hay que aclarar que el web scraping no está exento de polémica, debido a su legalidad y el derecho de recolectar los datos y su impacto en el rendimiento de los sitios web, sin embargo, para esta recolección se aseguró que el impacto fuera mínimo, además de considerarse publica cualquier información mostrada en una página web que no requiera registro o interacción humana adicional.

2.1 Herramientas utilizadas

Para la realización de esta investigación se usó la API de requests de Python, la cual se empleó para obtener solicitudes de tipo GET a los sitios de bolsa de trabajo, obteniendo toda la información contenida en la página, posteriormente se usó BeautifulSoup para la interpretación del código HTML recibido en el paso anterior.

De esta manera, se obtiene el código en HTML y se procesa con BeautifulSoup. Como nota adicional, el almacenamiento de los datos se llevó a cabo en una base de datos portable de SQLite 3, misma que posteriormente fue editada para eliminar registros con errores, datos incorrectos o incoherencias.

3 Análisis de los resultados

El presente apartado muestra los resultados de las operaciones de web scraping, principalmente la presencia e importancia que se le da a las habilidades blandas al momento de buscar un nuevo empleado.

La base de datos obtenida consta de tres tablas, la primera contiene las habilidades blandas que tienen un mayor número de menciones en artículos de investigación sobre las mismas, ejemplo de esto se encuentra en [11], donde se maneja la comunicación y el manejo del estrés como habilidades blandas esenciales.

La lista de habilidades manejadas (tanto en español e inglés) es la siguiente:

- Adaptabilidad
- Pensamiento crítico

- Pensamiento creativo
- Creatividad
- Trabajo en equipo
- Resolución de problemas
- Ética laboral
- Liderazgo
- Gestión del tiempo
- Pensamiento estratégico
- Comunicación
- Atención a los detalles
- Habilidades blandas (como termino genérico)
- Gestión del estrés
- Resiliencia

Las dos páginas que se usaron para la extracción de información fueron jobatus (<https://www.jobatus.mx/>) y occ mundial (<https://www.occ.com.mx/>), las cuales presentaban una compatibilidad con la API de requests y por tanto hicieron posible la recolección de datos.

3.1 Cuantificación de la información obtenida

La anteriormente mencionada base de datos se compone de 201 solicitudes de trabajo de dos sitios web, al mismo tiempo analizando su sección de descripción se encontraron 69 menciones de habilidades blandas en total, cabe destacar que esto no significa que haya 69 solicitudes de empleo con la mención de habilidades blandas, si no que en todas las 201 solicitudes solo hubo 69 menciones de las habilidades, pudiendo haber empleos pidiendo dos o más, lo cual a pesar de la limitada cantidad de información que disponemos el número de solicitudes pidiendo habilidades blandas es sorprendentemente bajo.

En total, de las 201 solicitudes analizadas solo 34 hicieron mención de habilidades blandas en sus requisitos o descripciones de empleo, lo cual no representa ni la mitad de las opciones que los aplicantes tienen, dando la impresión de que en el campo laboral no se usarán estas habilidades, sin embargo como se evidencia [12] en una fase posterior de reclutamiento (entrevista) las habilidades blandas son tomadas en cuenta, debido a que es precisamente en el campo laboral donde se mide que tan apto se es para comunicarse y trabajar en equipo.

Las solicitudes se centran más en los años de experiencia y las certificaciones requeridas, es decir que al conocimiento necesario se le da más prioridad que la capacidad de manejar el estrés y la adaptación a entornos de trabajo muy exigentes.

Alternativamente, queremos presentar la posibilidad que esta percepción se deba a que la publicación de la vacante no es el momento donde las empresas buscan que los solicitantes tengan este tipo de habilidades: Tal como se expresa en [12] y en [13], el uso de pruebas psicológicas es común en el proceso de reclutamiento, estas pruebas varían desde pruebas proyectivas hasta pruebas de aptitudes y habilidades, esta última, sería la candidata perfecta para ubicar la presencia de habilidades blandas en aplicantes, sin embargo esto requeriría más investigación.

A continuación, en la tabla 1, se presenta una muestra de las solicitudes que piden habilidades blandas y cuantas, así como un promedio de habilidades por solicitud.

Tabla 1. Muestra de los datos obtenidos, específicamente de Jobatus, se aprecia el número de habilidades blandas, así como el nombre de la vacante y el salario, en promedio, cuando existía la mención, se colocaban 2 habilidades blandas por solicitud.

OPERADOR PROGRAMADOR CNC PAGO POR HORA, EXPERIENCIA 2 AÑOS	Sueldo no mostrado por la compañía	1
PROGRAMADOR WEB EXPERIENCIA MÍNIMA NET Y JAVA	\$20000-\$30000	2
OPERADOR PROGRAMADOR CNC HAAS TM1, TM2	Sueldo no mostrado por la compañía	1
TORNO CNC OPERADOR Y PROGRAMADOR CNC	\$14000	1
OPERADOR	Sueldo no mostrado por compañía	1

PROGRAMADOR CNC PAGO POR HORA, EXPERIENCIA 2 AÑOS		
PROGRAMADOR/A JR	Sueldo no mostrado por compañía	1
PROGRAMADOR DE SOFTWARE JAVA BACKEND	Sueldo no mostrado por compañía	3
PROGRAMADOR SENIOR	Sueldo no mostrado por compañía	3
Tech Lead / Java Developer	\$35000-\$45000	1

3.2 Gráficas

Durante el análisis y revisión de los datos, se hizo evidente que es difícil darle una interpretación solo observando los valores obtenidos, por esa razón se decidió graficar algunos resultados para mejorar su comprensión.

Como se puede apreciar en la figura 1, las vacantes que solicitan habilidades blandas no sobrepasan el 20%, las limitaciones de esta investigación no permiten conocer si es el caso en otras áreas del conocimiento, sin embargo, es claro que al menos en el enfoque que tuvo esta actividad, los empleadores no tienen una fijación en reclutar gente con habilidades blandas o estas se evalúan en un ámbito diferente, por lo que sería necesario otro estudio para conocer donde se evalúa la presencia de estas habilidades (si se realiza).



Figura 1. Relación de solicitudes que piden habilidades blandas con los que no.

Centrándose más en las habilidades solicitadas, los resultados parecen favorecer a la comunicación y el trabajo en equipo (Véase figura 2) como habilidades predilectas que los reclutadores toman en consideración al momento de iniciar el proceso de reclutamiento (en caso de buscar gente con habilidades blandas desarrolladas), estas habilidades son muy solicitadas en el área de ingeniería en tecnologías de la información debido a que los empleos relacionados generalmente constan de equipos de trabajo, por lo que dichas habilidades son fundamentales.



Figura 2. Habilidades blandas más mencionadas en ofertas de empleo.

Como ultima gráfica, en la figura 3 se hace notar los salarios correspondientes a cada habilidad, sin embargo estos solo deben verse como la presentación de una hipótesis, debido a que para saber verdaderamente los rangos de salario de estas habilidades blandas, es necesaria una investigación a mayor escala con un numero de datos significativamente mayor, esto se debe que a veces las empresas no listan los salarios a los que aspiran los aplicantes, sin embargo, con los 30 datos de empleos con habilidades blandas que además mostraron sus salarios, se puede obtener la conclusión de que desarrollar la habilidad de la comunicación trae en promedio, un mejor salario que la más popular, el trabajo en equipo, nótese que generalmente para desarrollar una es necesaria la otra, pues ambos son necesarios para un buen ambiente laboral y personal.



Figura 3. Salario promedio por habilidad solicitada, sin contar puestos sin salario publico.

4 Conclusiones

Es importante mencionar que estas conclusiones están basadas en datos escasos, como ya hemos mencionado a lo largo de este trabajo, las conclusiones, aunque consideramos que son válidas y significativas, se necesitan más elementos y una investigación a mayor escala para poder llegar a tener un mayor impacto en nuestra sociedad con el propósito de desarrollar este tipo de habilidades que parecen pasar desapercibidas incluso para los empleadores, lo cual es indicativo de lo mencionado en [12], [13], es decir, las pruebas psicológicas usadas para la contratación de nuevos empleados.

Dichas pruebas son indicativas de la presencia de habilidades blandas, así como del sentido de responsabilidad. En caso de ser esto cierto, la verificación del dominio de estas habilidades seria al momento de la entrevista o posteriormente a la aplicación en línea o presencial del solicitante, explicando el aparente desinterés en solicitarlas en publicaciones en línea.

Referencias

- [1] Zhao, B. (2017). Web scraping. Encyclopedia of big data, 1 https://www.researchgate.net/profile/Bo-Zhao-3/publication/317177787_Web_Scraping/links/5c293f85a6fdccfc7073192f/Web-Scraping.pdf.
- [2] Lawson, R. (2015). Web scraping with Python. Packt Publishing Ltd. <https://ieeexplore.ieee.org/abstract/document/8822022/>
- [3] Marrero Sánchez, O., Mohamed Amar, R., & Xifra Triadú, J. (2018). Habilidades blandas: necesarias para la formación integral del estudiante universitario. REVISTA CIENTÍFICA ECOCIENCIA, 5, 1–18. <https://doi.org/10.21855/ecociencia.50.144>
- [4] Kouzis-Loukas, Dimitrios. Learning scrapy. Livery Place: Packt Publishing, 2016.
- [5] El Asikri, M. S., Knit, S., & Chaib, H. (2020). Using web scraping in a knowledge environment to build ontologies using Python and scrapy. European Journal of Molecular & Clinical Medicine, 7(03), 2020.
- [6] Richardson, L. (2007). Beautiful soup documentation.
- [7] Chapagain, A. (2019). Hands-On Web Scraping with Python: Perform advanced scraping operations using various Python libraries and tools such as Selenium, Regex, and others. Packt Publishing Ltd.
- [8] Christian Kempny, Patrick Brzoska. (2023). Anwendungskontexte von Web Scraping in der Versorgungsforschung - Nur für Web-Expert:innen? Oder eine Methode für alle Versorgungsforscher:innen!?

Zeitschrift für Evidenz Fortbildung und Qualität im Gesundheitswesen Volume 176. 61-64.
<https://doi.org/10.1016/j.zefq.2022.11.010>

(<https://www.sciencedirect.com/science/article/pii/S1865921722002288>)

- [9] Wang, Z., Zeng, J., Ran, H., Meng, W., Zhou, S., Davies, A. B., & Liu, C. (2023). Monitoring the online ant trade reveals high biological invasion risk. *Biological Conservation*, 282. <https://doi.bibliotecabuap.elogim.com/10.1016/j.biocon.2023.110038>.
- [10] Macias, P., Stelmasiak, D., & Szafranek, K. (2023). Nowcasting food inflation with a massive amount of online prices. *International Journal of Forecasting*, 39(2), 809–826. <https://doi.bibliotecabuap.elogim.com/10.1016/j.ijforecast.2022.02.007>
- [11] Espinoza Mina, Marcos Antonio, & Gallegos Barzola, Doris. (2020). Habilidades blandas en la educación y la empresa: Mapeo Sistemático. *Revista Científica UISRAEL*, 7(2), 39-56. <https://doi.org/10.35290/rcui.v7n2.2020.245>.
- [12] Vidal, A. (2023). Conoce las principales fases del proceso de selección. *TalentFY*. <https://ytalentfy.com/fases-proceso-de-seleccion/>
- [13] Sánchez, D. (2018). Técnicas de selección utilizadas para la gestión del talento humano en tiendas por departamento del municipio Maracaibo. *CICAG: Revista del Centro de Investigación de Ciencias Administrativas y Gerenciales*, 15(2), 95-107.