

USO DE API PARA EL ANÁLISIS DE CITAS EN LAS BIBLIOTECAS DE LA FES CUAUTILÁN UNAM

José Luis Garza-Rivera^{1*}, Angélica Espinoza-Godínez², Jessica Annabel Páez-Arancibia¹ y Margarita Micaela Zapata-Guerrero¹

¹*Coordinación de Bibliotecas y Hemerotecas*

²*Centro de Tecnologías en Cómputo y Comunicación
Facultad de Estudios Superiores Cuautitlán.*

*jlgr@unam.mx

Resumen

Uno de los requerimientos habituales, de la comunidad académica y de posgrado, al personal académico de las bibliotecas y centros documentales, es el análisis de citas de sus publicaciones. La petición se deriva de las solicitudes de estos análisis, por parte de los consejos de ciencia, tecnología y humanidades, quienes utilizan esta información como un elemento para la medición de la productividad académica de aquellos que desean acceder a los programas de apoyo a la investigación y sistemas de becas. Para el análisis de citas, el personal bibliotecario cuenta con la formación y las herramientas que permiten obtener la información necesaria, sin embargo, cuando la persona de quien se hace el análisis de citas cuenta con un número considerable de publicaciones o citas, el proceso se vuelve complejo y puede requerir de muchas validaciones, limpieza de datos y tiempo. El objetivo de este trabajo es complementar los procedimientos habituales, que se realizan mediante las plataformas de los proveedores de información para la búsqueda de las publicaciones y sus citas, así como gestores de referencia, con el uso de interfaces de desarrollo de aplicaciones (API). Para este desarrollo, la plataforma usada junto con su portal y API fue Scopus de Elsevier, mediante el lenguaje de programación Python y las bibliotecas de uso libre PyScopus y PyBibliometrics. El manejo de los resultados se efectuó con el gestor de referencias Mendeley, también de Elsevier, en sus versiones de cliente para PC y Web. Las aplicaciones resultantes

permiten obtener los datos de forma más rápida y precisa, lo que mejora el procedimiento y facilita la verificación de la información obtenida.

Palabras clave: Programación de aplicaciones, gestores de referencias, métricas de publicación, Python, bibliotecas universitarias.

Introducción

El análisis de citas es una herramienta que tiene varios usos, uno de ellos es como elemento en la medición de la productividad académica de quienes escriben artículos científicos, las instituciones donde se desenvuelven o los campos del conocimiento en los que trabajan, incluso se puede utilizar para estimar la productividad académica por país.

El análisis de citas, como parte de la bibliometría que es la aplicación de métodos cuantitativos para estudiar los recursos de información, se refiere al estudio de las frecuencias y patrón de las citas en libros y revistas (Richard & Rachel, 2020).

Cuando los autores citan un trabajo dentro del propio, además de indicar la autoría de una idea, concepto o texto creado por alguien más, establecen que el texto citado influyó en mayor o menor medida sobre el documento que está presentando (Cañedo 1999; Miguel *et al.*, 2007).

El que los trabajos de una persona sean referenciados por otras personas autoras, eliminando las autocitas y tomando las consideraciones necesarias sobre las citas en trabajos de los coautores, sirve como un indicador de la influencia o impacto en el área de estudio. Rosa Sancho lo refiere como Número de Citas recibidas (de otras publicaciones posteriores) (Sancho, 1990).

Es común que los consejos de ciencia, tecnología y humanidades; ya sean nacionales, regionales o locales, tomen como un elemento de decisión, para la evaluación de

candidatos, las citas recibidas en las publicaciones consideradas dentro de sus listados o bases de datos de índices de publicaciones académicas.

Dentro de las bases de datos con índices de publicaciones académicas disponibles para el análisis de citas, entre las que tienen mayor relevancia, podemos encontrar a Scopus de Elsevier (Elsevier, 2023), Web of Science de Clarivate (Clarivate, 2023b) y Google Académico (Google LLC, 2023). Estos tres recursos son los considerados en México por el Consejo Nacional de Humanidades, Ciencias y Tecnologías (CONAHCYT) (CONAHCYT, 2020), al momento de realizar esta investigación, con la distinción de autocitas, citas externas (tipo A) y co-citas (tipo B).

Las tres bases de datos mencionadas en el párrafo anterior brindan la posibilidad de trabajar en línea mediante portales o bien utilizando interfaces de desarrollo de aplicaciones que pueden integrarse a diversos lenguajes de programación como Python. En el caso de Web of Science y Scopus, las bases ofrecen opciones de acceso limitadas para el público en general y usuarios sin licencia; y opciones extendidas para usuarios con acceso por suscripción (Clarivate, 2023a; Elsevier B.V., 2023), siendo esta última opción la disponible para las bibliotecas del Sistema Bibliotecario de la UNAM. Para Google Académico, el acceso es gratuito y se pueden obtener bibliotecas de acceso desarrolladas por terceros, algunas de ellas de paga (Bettenbuk, 2021).

Para este desarrollo se eligió, como punto de partida, el uso del API de Scopus debido a que es la primera base de datos empleada para el análisis de citas, contiene más registros y permite la conformación directa de perfiles de autores. Una vez que generamos el reporte de Scopus, lo complementamos con la información no redundante obtenida de las otras dos bases de datos, Web of Science y Google Académico.

Uno de los lenguajes de programación más utilizados para la ciencia de datos y procesamiento de información documental es Python (Python Software Foundation, 2023), este lenguaje, mediante el uso de bibliotecas incorporadas o desarrolladas por terceros, permite la creación de proyectos informáticos, incluidos los de bibliometría.

Objetivo

Complementar los procedimientos habituales, que se realizan mediante las plataformas de los proveedores de información para la búsqueda de las publicaciones y sus citas, así como gestores de referencia con el uso de interfaces de desarrollo de aplicaciones (API).

Análisis de citas realizados

Los análisis de citas efectuados se componen de un listado de referencias de los artículos donde la persona solicitante es autora, seguido por cada una de las referencias de sus citas, clasificadas en tipo A, B y autocitas. Se incluye un conteo de cada tipo. En los casos en que se tienen pocas publicaciones o las publicaciones cuentan con pocas citas, el trabajo que puede hacerse de manera rápida en la página de Scopus, vía la Biblioteca Digital de la UNAM (Dirección General de Bibliotecas y Servicios Digitales de Información, 2023) o directamente en el sitio de Scopus en Elsevier, aunque este último método no permite el acceso a algunas características.

El sitio web de Scopus cuenta con perfiles de autor, en el que tratan de agrupar, de forma automática, las publicaciones de la misma persona. Estos perfiles se asocian con un *Scopus Author Identifier*. Los elementos que generan agrupaciones erróneas son la existencia de homonimias, el uso de rúbricas distintas por parte de los autores, la falta de uso de identificadores como el ORDIC (ORCID, 2023), así como el cambio de adscripciones laborales o de investigación.

Para evitar omitir publicaciones o incluir publicaciones de otros autores, como en el caso de las homonimias, se pide un listado de sus publicaciones a las personas que requieren el análisis. Un primer paso es verificar cuáles de esas publicaciones se encuentran indexadas en Scopus, y en caso de que abarquen más de un *Scopus Author Identifier*, solicitamos a Elsevier que los agrupe en un mismo identificador. Esto facilita los procesos posteriores para la identificación de autocitas y las citas tipo A y B.

En la Figura 1 se observa, de manera resumida, el procedimiento para obtener el reporte del análisis de citas en Scopus.



La generación del reporte final mediante este procedimiento implica tener abierto de manera simultánea la página Web de la base de datos de Scopus, la versión en línea y de escritorio del gestor de referencias, en este caso Mendeley; así como el procesador de palabras a emplear. Aunque el sitio web permite filtrar y preseleccionar cierta información, se requiere seleccionar y verificar constantemente datos que pasan de una aplicación a otra y de una correcta sincronización en el paso de datos, ya sea por exportación directa o desde el portapapeles del sistema operativo.

Recursos de programación de Scopus para el acceso programático a la búsqueda y extracción de información

Elsevier proporciona un conjunto de herramientas para el desarrollo de aplicaciones, que pueden obtenerse y explorarse desde su portal (Elsevier B.V., 2023), entre ellas podemos destacar la biblioteca `elsapy` para Python, llaves de acceso (*API Keys*) que permiten el acceso programático a los datos (algunos recursos requieren una solicitud por escrito a Elsevier para obtener o activar llaves específicas) y un sitio interactivo de pruebas de API para generar y probar segmentos de código. Por parte de terceros, existen bibliotecas para Python como `PyScopus` y `PyBibliometrics` (Zuo *et al.*, 2017a, 2017b; Rose & Kitchin, 2019).

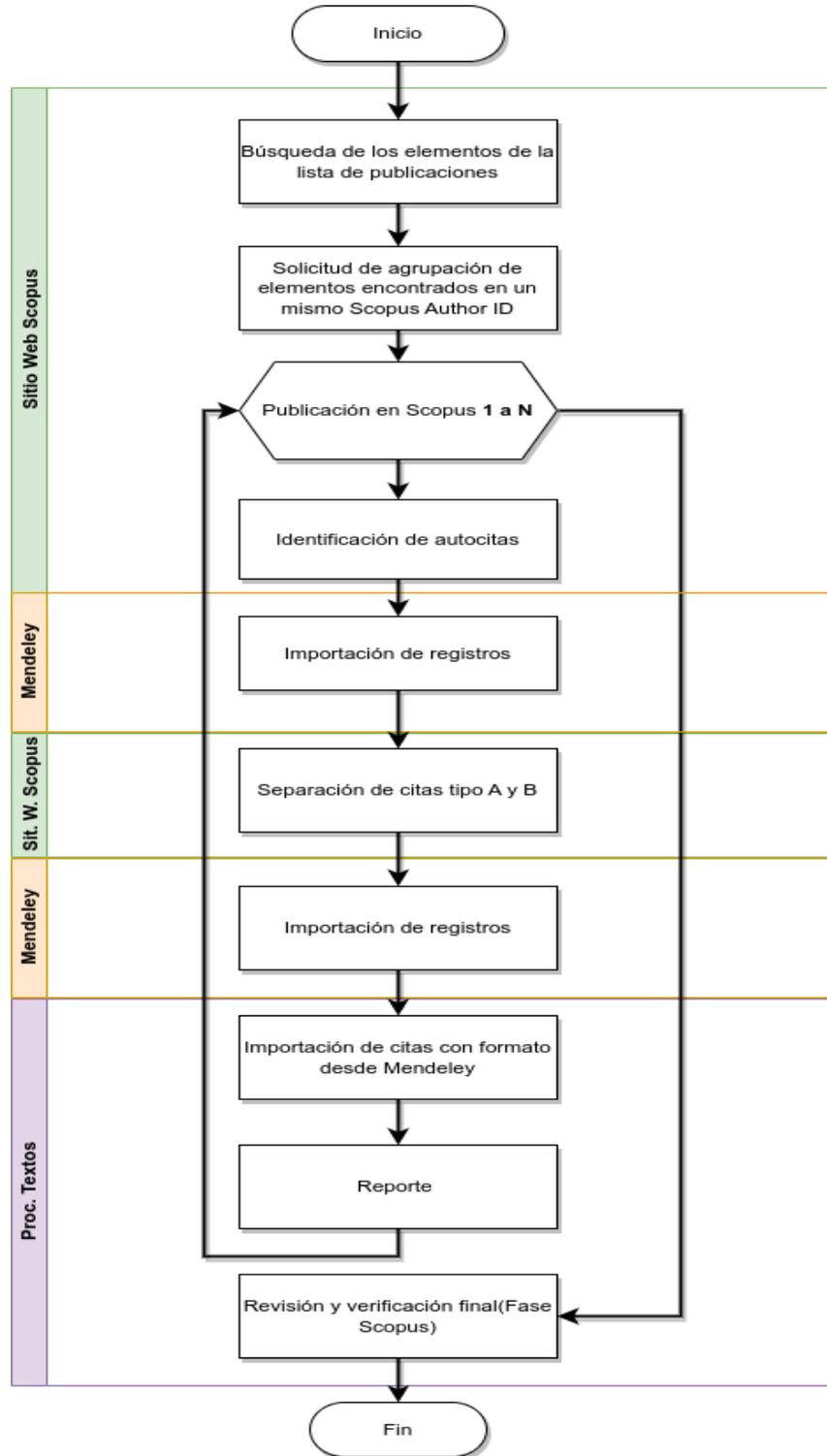


Figura 1. Procedimiento con el portal Scopus y Mendeley (Fuente: elaboración propia).



Optimización del procedimiento con Python y Bibliotecas PyScopus y PyBibliometrics

El proceso descrito en la Figura 1 se automatizó empleando las funciones `ElsAuthor`, `AbsDoc`, `ElsSearch`, `REFEID` de la biblioteca `elsapy`. En el caso de `PyBibliometrics` se usaron funciones como `AuthorRetrieval`, `ScopusSearch`, `AbstractRetrieval`, `citedby_link`. La biblioteca `PyScopus` fue empleada para la obtención y validación de diversos datos.

Una vez que se generan las citas, estas se insertan en documentos mediante el empleo de las bibliotecas Python `openpyxl` y `docx`.

El procedimiento final se ilustra en la Figura 2, donde se identifica la etapa de sitio web Scopus en color verde para realizar los procesos iniciales de la lista de publicaciones. En la etapa Python en color naranja son ejecutadas las API de los procesos de análisis de citas, desde 1 hasta N, es decir, desde la primera publicación hasta la última de la lista, generando el reporte correspondiente y, en la etapa de procesamiento de textos en color morado, se revisan y validan los registros del reporte.

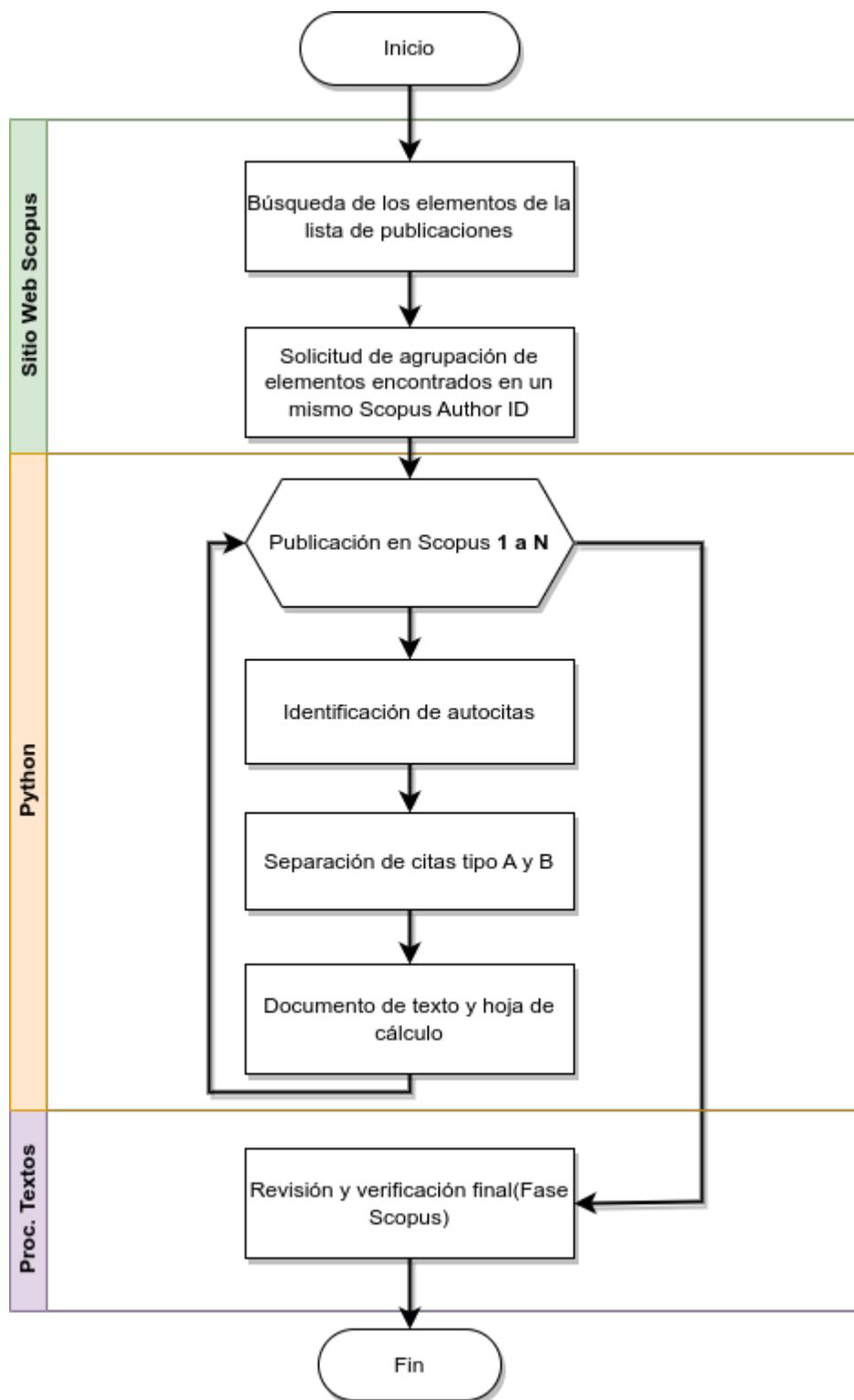


Figura 2. Procedimiento con las API de Python.



Conclusión

Trabajar con Python, así como con las API de Elsevier y de terceros, permite generar de manera más rápida y desatendida los reportes de análisis de citas. El proceso de verificación y limpieza de resultados también se simplifica. La confirmación de la información es más sencilla y no requiere aplicaciones concurrentes como son el gestor de referencias, el navegador en el sitio del Elsevier y el programa de procesamiento de textos. Como un trabajo a futuro, se integrarán los otros dos servicios de bases de datos consultadas, Web of Science y Google Académico. Las aplicaciones desarrolladas impactan de forma positiva a los servicios de análisis de citas proporcionados por las bibliotecas de la Facultad de Estudios Superiores Cuautitlán a su comunidad académica y los estudiantes de posgrado.

Agradecimientos

Se agradece el apoyo del programa UNAM-DGAPA-PAPIME con los proyectos PE207422, PE202323, PE103023 y PE103223. Asimismo, al programa UNAM-DGAPA-PAPIIT con el proyecto IA102323. Agradecemos también el apoyo de Katherine Ruth García de Customer Experience Champion ELSEVIER | Data for Research and Discovery, por el soporte técnico, activación de llaves y liberación de acceso a los elementos restringidos de la plataforma de Scopus.

Referencias

Bettenbuk, Z. (2021). *5 Best Google Scholar APIs and Proxies for 2023 - ScraperAPI*. <https://www.scraperaapi.com/blog/best-google-scholar-apis-proxies/>

Cañedo, A. R. (1999). Los análisis de citas en la evaluación de los trabajos científicos y las publicaciones seriadas. *Acimed*, 7(1): 30-39.

Clarivate. (2023a). *Clarivate Developer Portal - Web of Science API Expanded*. <https://developer.clarivate.com/apis/wos>

Clarivate. (2023b). *Web of Science*. <https://access.clarivate.com/login?app=wpp>

CONAHCYT. (2020). *PREGUNTAS FRECUENTES Convocatoria para Ingreso o Permanencia en el Sistema Nacional de Investigadores*. <https://conahcyt.mx/wp->



content/uploads/sni/preguntas_frecuentes_sni/faq2020.pdf

Dirección General de Bibliotecas y Servicios Digitales de Información. (2023). *Biblioteca Digital UNAM - Ver todas las Bases de Datos*.

<https://www.bidi.unam.mx/index.php/colecciones-digitales/bases-de-datos/ver-todos-los-recursos>

Elsevier. (2023). *Scopus - Document search*. <https://www.scopus.com>

Elsevier B.V. (2023). *Elsevier Developer Portal*. <https://dev.elsevier.com/>

Google LLC. (2023). *Google Académico*. <https://scholar.google.es/schhp?hl=es>

Miguel, S., Moya, A. F. & Herrero, S. V. (2007). El análisis de co-citas como método de investigación en Bibliotecología y Ciencia de la Información. *Investigación Bibliotecológica: archivonomía, bibliotecología e información*, 21(43): 139-155.

<https://doi.org/10.22201/iibi.0187358xp.2007.43.4129>

ORCID. (2023). *ORCID*. <https://orcid.org/>

Python Software Foundation. (2023). *Welcome to Python.org*. <https://www.python.org/>

Richard, E. R. & Rachel, G. R. (2020). *Foundations of Library and Information Science.: Vol. Fifth edit.* ALA Neal-Schuman.

Rose, M. E. & Kitchin, J. R. (2019). *pybliometrics: Scriptable bibliometrics using a Python interface to Scopus*. *SoftwareX*. <https://doi.org/10.1016/j.softx.2019.100263>

Sancho, R. (1990). Indicadores bibliométricos utilizados en la evaluación de la ciencia y la tecnología. Revisión bibliográfica. *Revista española de documentación científica*, 13(3-4): 842-865.

Zuo, Z., Zhao, K. & Eichmann, D. (2017a). *PyScopus: Get Started*. <http://zhiyuzuo.github.io/python-scopus/>

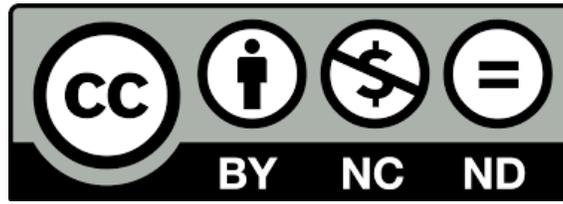
Zuo, Z., Zhao, K. & Eichmann, D. (2017b). The state and evolution of U.S. iSchools: From talent acquisitions to research outcome. *Journal of the Association for Information Science and Technology*, 68(5): 1266-1277. <https://doi.org/10.1002/ASI.23751>



D. R. © UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO.

Excepto donde se indique lo contrario esta obra está bajo una licencia Creative Commons Atribución No comercial, No derivada, 4.0 Internacional (CC BY NC ND 4.0 INTERNACIONAL).

<https://creativecommons.org/licenses/by-nc-nd/4.0/deed.es>



ENTIDAD EDITORA

Facultad de Estudios Superiores Cuautitlán.

Av. Universidad 3000, Universidad Nacional Autónoma de México, C.U., Delegación Coyoacán, C.P. 04510, Ciudad de México.

FORMA SUGERIDA DE CITAR:

Garza-Rivera, J. L., Espinoza-Godínez, A., Páez-Arancibia, J. A., y Zapata-Guerrero, M. M. (2023). Uso de API para el análisis de citas en las bibliotecas de la FES Cuautitlán UNAM. *MEMORIAS DEL CONGRESO NACIONAL DE TECNOLOGÍA (CONATEC)*, Año 6, No. 6, septiembre 2023 - agosto 2024. Facultad de Estudios Superiores Cuautitlán. UNAM.

https://tecnicosacademicos.cuautitlan.unam.mx/CongresoTA/memorias2023/Mem2023_Paper06-E.html