

SIESTA. Una aplicación Web para obtener indicadores bibliométricos en ciencia y tecnología

SIESTA. A Web application for obtaining science and technological bibliometric indicators

Madera-Jaramillo, MJ
CUIB, UNAM
Torre II de Humanidades
Pisos 11 al 13
Circuito Interior
Cd. Universitaria, D.F. 04510
Tel. +52 (55)5623-0333
e-mail: maji@servidor.unam.mx

López Gaona, Amparo
Fac. de Ciencias, UNAM
Circuito Interior
Cd. Universitaria, D.F. 04510
Tel. +52 (55)5622-4866
e-mail: alg@ciencias.unam.mx

Ruiz León, Alejandro
IIMAS, UNAM
Circuito Interior
Cd. Universitaria, D.F. 04510
Tel. +52 (55)5622-6230
e-mail:
rarnulfo@servidor.unam.mx

Resumen

La tarea de evaluar la producción científica ha adquirido gran importancia porque, a partir de ella, se pueden definir lineamientos para la asignación de recursos económicos de forma más adecuada. A finales de la década de los 90 las aplicaciones *Web* atrajeron la atención de los encargados de analizar y evaluar la producción científica pues este tipo de aplicaciones permite que la información difundida a través de *Internet* se actualice en forma constante y pueda ser utilizada por una gran cantidad de individuos. El objetivo del presente trabajo es presentar la aplicación *Web* SIESTA, se trata de una herramienta que proporciona los mecanismos para normalizar los datos de artículos mexicanos tomados de la base de datos *Current Contents*. A partir de estos datos, a través de su interfaz de consulta, es posible obtener indicadores bibliométricos por área geográfica, disciplinas y organizaciones, y conocer rápida y fácilmente la producción científica mexicana con visibilidad internacional.

Palabras clave: Aplicación web, Indicadores bibliométricos, Visibilidad internacional mexicana

Introducción

En la actualidad la mayoría de las empresas e instituciones académicas están haciendo uso del avance tecnológico para difundir información de forma rápida, fácil y a bajo costo. Esto se debe a que la *Internet* proporciona los mecanismos que apoyan el proceso de difusión. La creciente popularidad de la *Internet*, gracias al surgimiento de la *World Wide Web* (WWW o Web), generó la necesidad de

desarrollar aplicaciones *Web*. Una de las principales ventajas de este tipo de aplicaciones es permitir que la información difundida a través de ellas se actualice en forma constante y pueda ser utilizada por una gran cantidad de individuos. A finales de los años 90 las aplicaciones *Web* atrajeron la atención de los encargados de analizar y evaluar la producción científica. Esta necesidad de evaluar la producción científica ha adquirido gran importancia debido porque a partir ella se pueden definir lineamientos para la asignación de recursos económicos de forma más adecuada. El análisis del desarrollo de la ciencia se inició a principios del siglo XX, Sancho [1] asevera que "...algunos de los primeros trabajos en bibliometría fueron resultado de una curiosidad innata por entender el desarrollo científico", y que la evaluación de la producción científica se hace a través de indicadores bibliométricos porque éstos proporcionan información que permite conocer los avances y el impacto de la ciencia.

La necesidad de un grupo italiano que se encontraba realizando un análisis sobre la producción científica de dicho país tomado como fuente de datos las ediciones *Life sciences(LS)*, *Agriculture, Biology & Environmental sciences(ABES)*, *Engineering, Computing & Technology(ECT)* y *Physical, Chemical & Earth sciences(PCES)* de la base de datos *Current Contents(CC)*, hizo que se desarrollara la aplicación *Web* SIESTA (*Software Interface for the Scientific Technological Activity*), la cual se usa para obtener indicadores bibliométricos en ciencia y tecnología italianos por área geográfica, por organizaciones y por disciplinas. Además es posible obtener la colaboración de Italia con otros países [2]. Algunas de las instituciones que han desarrollado aplicaciones que permiten obtener indicadores de ciencia y tecnología a través de la *Web* son: la Universidad de Sussex en Inglaterra [3] y el Centro de Información y Documentación Científica (CINDOC) en España [5].

Objetivo

El objetivo del presente trabajo es presentar la adaptación de la aplicación SIESTA, utilizando la misma fuente de datos, para obtener indicadores

bibliométricos mexicanos con el propósito de determinar la producción científica por estado, ciudad, organizaciones y disciplinas. No obstante que la fuente seleccionada para obtener indicadores en el caso de Italia proporciona una cobertura mínima de la producción científica mexicana, la adaptación de SIESTA al caso México tendrá gran utilidad y valor para los encargados de tomar las decisiones en política científica mexicana ya que corresponde a la producción científica nacional más visible a nivel internacional, la denominada *mainstream*.

Materiales y métodos

Puesto que, en cualquier campo, las publicaciones son los principales resultados de la actividad científica es importante mostrar el proceso de generación del conocimiento científico, cómo éste es difundido entre la comunidad científica y posteriormente utilizado para realizar análisis de desarrollo y productividad científica. La investigación se inicia con un dato, una evidencia o una información que es difundida a través de los medios de comunicación ya sean formales o informales y por medio de éstos llega a ser publicada en revistas, que es uno de los resultados esperados en la investigación científica (artículos). La publicación en estos medios formales de comunicación fomenta el desarrollo de bases de datos a los que se puede acceder para obtener indicadores bibliométricos.

La figura 1 muestra un diagrama de un modelo de la comunicación científica citado por Hurd [5], en el que los resultados de la investigación científica, antes de su comunicación formal a través de su publicación como artículos en revistas científicas, son difundidos frecuentemente en seminarios y conferencias. Los datos de los artículos son recogidos en primera instancia por el servicio de alerta de *Current Contents* para luego ser indexados en las bases de datos bibliográficas. La base de datos *Current Contents* es generada por el *Institute for Scientific Information* (ISI) de Filadelfia, y provee acceso a la información bibliográfica más completa sobre artículos, editoriales, cartas, comentarios, resúmenes de reuniones y reseñas de libros. Esta base de datos es actualizada semanalmente,

lo que ayuda a los usuarios a mantenerse actualizados y les proporciona un panorama completo de la investigación global al día.

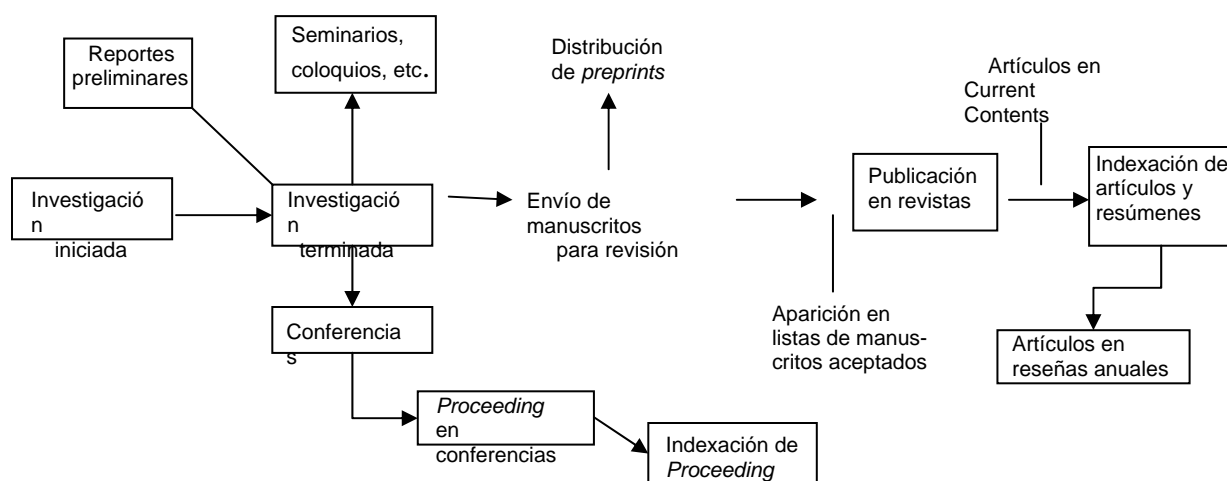


Figura 1. Modelo de comunicación científica Garvey-Griffith

La importancia de utilizar la base de datos *Current Contents* para obtener indicadores bibliométricos radica en el hecho de su frecuente actualización, así el análisis del desarrollo de la ciencia realizado con estos indicadores está siempre al día; en cambio el análisis que se hace utilizando las otras bases de datos bibliográficas es retrospectivo ya que éstas son generadas posteriormente a que los datos sean almacenados en *Current Contents*.

Características funcionales que proporciona esta aplicación Web

Puesto que se van a obtener indicadores bibliométricos en ciencia y tecnología para conocer la producción científica mexicana más visible a nivel internacional, la fuente de datos utilizada es el *Current Contents*, y al igual que en el caso italiano se procesan artículos de las ediciones ABES, ETC, LS y PCES.

Como un primer paso para iniciar la normalización de la información extraída de *Current Contents* (en adelante CC) se creó un tesoro con los nombres normalizados de los países, tomando una lista proporcionada por ISO-3166 Maintenance Agency (ISO-3166/MA). También, se creó un tesoro que tomó los

nombres de las ciudades y estados mexicanos tomados del Instituto Nacional de Estadística, Geografía e Informática.

Los datos extraídos de la base de datos CC son los siguientes: Accession number, ISSN, CC categories, Institutions.

He aquí las características funcionales con que cuenta la aplicación SIESTA:

1. Interfaz gráfica para la *Web* que permite obtener indicadores bibliométricos a nivel estado, ciudad y organización, así como indicadores que muestran la colaboración científica de México con otros países, desde cualquier lugar físico que tenga conexión a *Internet*.
2. Procesamiento de datos extraídos de CC:
 - Cada registro extraído de CC contiene el nombre del país y las instituciones que participaron en la producción del artículo, y si el nombre de un país que no sea México aparece más de una vez solamente se contabiliza una colaboración. La codificación del país se hace de acuerdo con el tesauro y los nombres de los países no codificados se utilizan para crear un tesauro de sinónimos de países y así proceder con el proceso de codificación.
 - Para las instituciones cuyo país de origen es México, se realiza una codificación que este de acuerdo con el tesauro generado con la lista de la INEGI. Al igual que en el caso de países, se genera un tesauro de sinónimos de ciudades y posteriormente éste es utilizado para codificar los nombres de las ciudades no normalizadas.
 - Se crea un tesauro de organizaciones mexicanas (UNAM, UAM, IMSS, etcétera.), se almacenan abreviaciones y acrónimos de cada una de las organizaciones y se realizar la codificación de los nombres de las “instituciones” mexicanas extraídas de la base de datos CC.

Es importante recalcar que debido a la complejidad que implica la normalización de los datos solamente se normalizaron los nombres de las ciudades e instituciones mexicanas.

Resultados

Con la aplicación SIESTA el proceso de normalización de los datos de los artículos extraídos de la base de datos CC es fácil y rápida, y la obtención de indicadores bibliométricos se puede hacer en el momento en que el usuario los necesite.

La figura 2 muestra un panorama general del proceso utilizado por la aplicación para obtener indicadores bibliométricos. Se extraen los artículos mexicanos correspondientes a las ediciones ABES, LS, ECT y PCES, se crea un conjunto intermedio de datos de los cuales son extraídos los datos de los artículos y de las instituciones participantes, y se generan tesauros de sinónimos de ciudades, países y organizaciones. Con los tesauros y los datos de los artículos se obtienen indicadores bibliométricos por número y distribución de las publicaciones en términos de áreas geográficas, disciplinas y organizaciones.

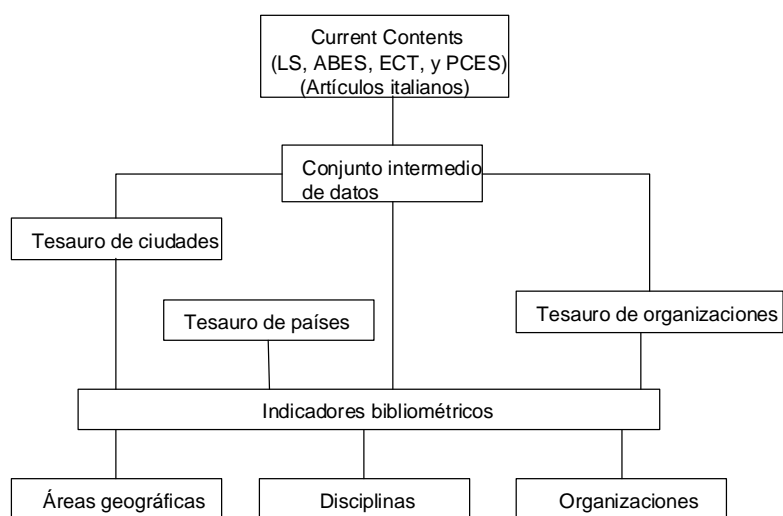


Figura 2. Esquema general de la aplicación SIESTA

En la figura 3 se muestra un ejemplo de una consulta hecha a través de la interfaz de consulta de SIESTA. En la figura 3a el usuario selecciona consultar artículos de la edición ABES, en la figura 3b se muestra la página donde se deberá seleccionar el área geográfica, en nuestro caso se desea conocer la colaboración

de México con otros países, así que se selecciona la opción País. Una vez que el usuario elige el área geográfica, se genera una página en la cual se deberá seleccionar el criterio de consulta, ésta se muestra en la figura 3c. En este ejemplo se quiere saber la producción científica de México con USA o Canada o United Kingdom sobre todas las disciplinas y años de la edición ABES, y por lo tanto se seleccionan estos criterios de consulta, finalmente, en la figura 3d se muestra la página de resultados que contiene el número de artículos en los que México ha colaborado con cada uno de los tres países antes mencionados.

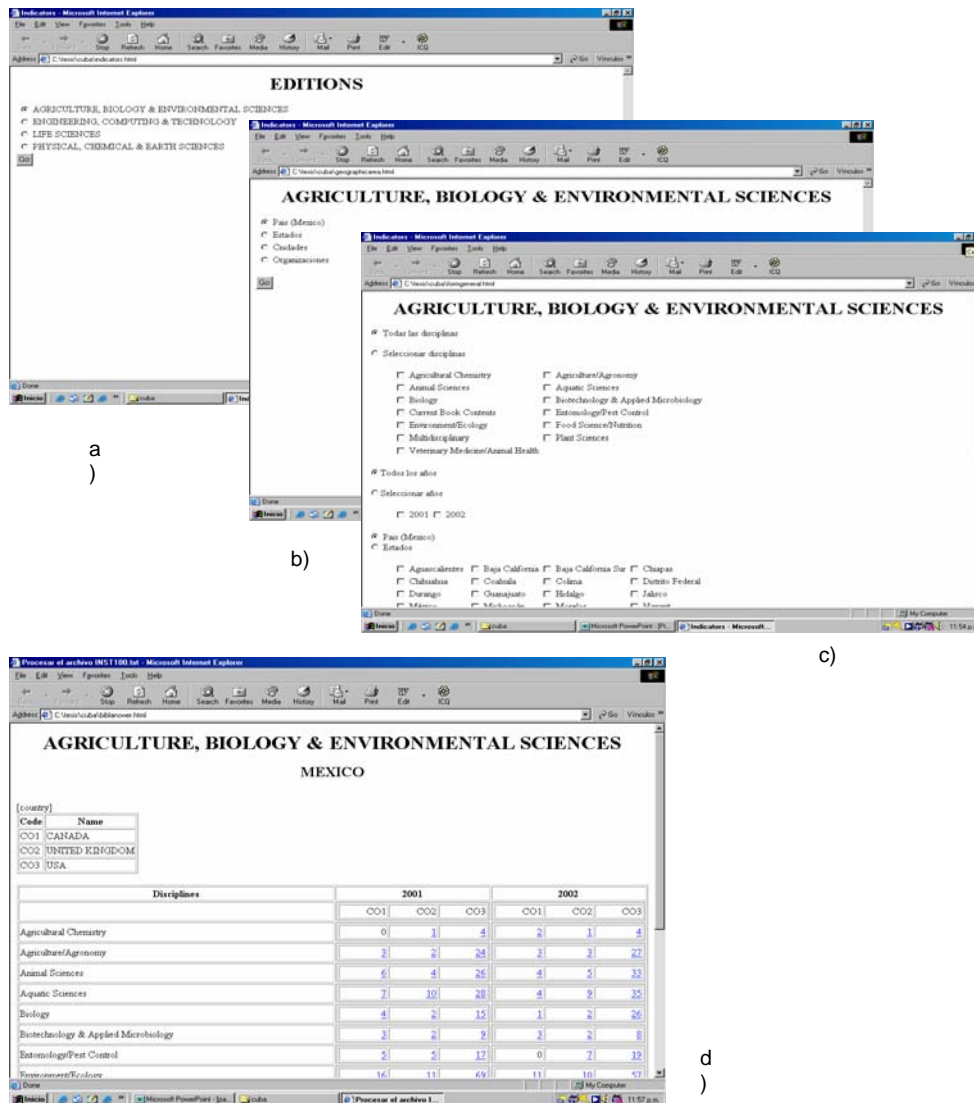


Figura 3. Ejemplo de consulta a la aplicación SIESTA

Con los resultados que obtenidos a través de la interfaz de consulta de SIESTA es posible hacer estudios como los que se muestran en la siguiente sección.

Representación gráfica de colaboraciones

En esta sección se presenta la aplicación de técnicas empleadas en análisis de redes sociales [6], en particular lo referente a la parte de la visualización de características estructurales, en el análisis de la colaboración científica.

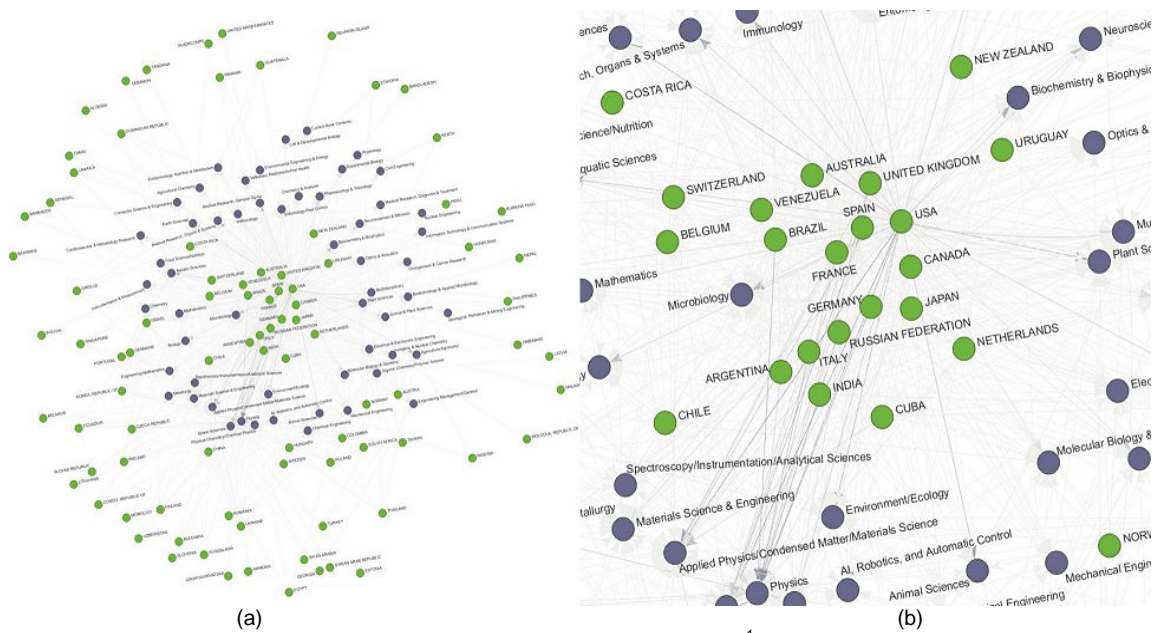


Figura 4. Colaboración por disciplina¹.
país (nodos verdes), disciplinas(nodos azules)

En la figura 4a se muestra la representación gráfica de la colaboración por disciplina de las cuatro ediciones del CC para el año 2002, utilizando un método para modelar centralidad en análisis de sociales [7]. En la parte central de la gráfica se localizan los países con los que se tiene mayor colaboración, y hacia la periferia se localizan los de menos colaboración; los países están relacionados con las disciplinas de tal manera que entre mayor colaboración se tenga con una disciplina, menor será su separación física, esto permite distinguir grupos de países con los que se tiene una colaboración similar. Si analizamos las disciplinas en las que se colaboró con los países que se localizan al centro de la gráfica

¹ Para la elaboración de las gráficas se utilizó PAJEK, <http://vlado.fmf.uni-lj.si/pub/networks/pajek/default.htm>

(USA, SPAIN, FRANCE, UNITED KINGDOM, GERMANY, CANADA, JAPAN) (Figura 4b) encontramos que MÉXICO tuvo una colaboración similar por disciplina con estos países.

Conclusiones

Sin duda alguna es importante para quienes toman las decisiones contar con una herramienta que los apoye para obtener indicadores bibliométricos en ciencia y tecnología, y para evaluar de una forma rápida y fácil la producción científica mexicana que tiene visibilidad internacional. SIESTA es una herramienta de apoyo para la obtención de indicadores bibliométricos que puede ser usada para fines académicos de forma gratuita, ya que se desarrolló con fondos proporcionados por una institución académica, pero debido a que la fuente de datos utilizada es de ISI éstos deben emplearse siguiendo las normas marcadas por él en cuanto a derechos de autor.

A SIESTA se le pueden agregar nuevos módulos para obtener otros indicadores, como por ejemplo el factor de impacto de las revistas, el cual es una herramienta muy útil para determinar el valor de las publicaciones, desde el punto de vista de la frecuencia con la cual que se citan sus artículos en trabajos posteriores. SIESTA se encuentra disponible en <http://alcatraz.fciencias.unam.mx/~mji/siesta/>.

Referencias bibliográficas

1. Sancho, R. Indicadores bibliométricos utilizados en la evaluación de la ciencia y la tecnología. Revisión bibliográfica. *Revista Española de Documentación Científica*, 13:842-865, 1990.
2. Madera-Jaramillo, MJ. Tesis de maestría: *Una aplicación Web para la obtención de indicadores bibliométricos en ciencia y tecnología*. El autor, Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas, México, D.F., 2001.
3. Katz, J.S. y Hicks, D. Desktop scientometrics. *Scientometrics*, 38(1):141-153, 1997. y Katz, J.S. y Hicks, D. Indicators for systems of innovation, a bibliometrics-based approach; [en línea]. *IDEA paper series* 12, Reino Unido: Step Group, <http://www.sussex.ac.uk/Users/sylvank/pub/Idea12.pdf>, [Consulta: Marzo de 2001]. 1998.
4. Fernández, M.T. INDICYT: Indicadores de ciencia y tecnología en red. Cuadernos de Indicios. Indicadores bibliométricos en Iberoamérica. Red

- de Indicadores de Ciencia y Tecnología-RICYT, (1):101-109, 2001 y Fernández, M.T. et al. INDICYT:Science and Technology indicators in Spain: development of an application for interactive search on the Internet. Research Evaluation, 10(2):83-88, 2001.
5. Hurd, JM. The transformation of scientific communication: A model for 2020. Journal of the American Society for Information Science, 51(14):1279-1283, 2000.
 6. Wasserman, S. y K. Faust. Social Network Analysis: methods and applications, Cambridge: Cambridge University Press. New York. 1994.
 7. V.Batagelj, A. Mrvar. Pajek – Program for Large Network Analysis, Connections 21(2):47-57, 1998. asignar