

Analysis and data visualization in bibliometric studies

Joel Alhuay-Quispe^(a), Alonso Estrada-Cuzcano^(b),
Lourdes Bautista-Ynofuente^(c)

a) Universidad Privada San Juan Bautista, Facultad de Ingeniería. Lima, Perú, <https://orcid.org/0000-0002-1903-4687>

b) Universidad Nacional Mayor de San Marcos. Lima, Perú, <https://orcid.org/0000-0001-5039-1108>

c) Universidad Nacional Mayor de San Marcos. Lima, Perú, <https://orcid.org/0000-0002-0277-489X>

Contact: Joel Alhuay-Quispe, joel.alhuay@upsjb.edu.pe; Alonso Estrada-Cuzcano, mestradac@unmsm.edu.pe;
Lourdes Bautista-Ynofuente, lourdes.bautista1@unmsm.edu.pe

Received: 14 July 2021; **Accepted:** 17 January 2022; **First Published:** 15 May 2022

ABSTRACT

In recent years, academics with different professional formation other than Information Sciences have been involved in research related to metric studies using approaches and techniques based on bibliometrics; however, these professionals do not have knowledge and necessary competences for conduct metric information studies in a rigorous and methodical way. This article aims to find what are methods and tools used by bibliometrics researchers through a descriptive and textual analysis, using methods such as word co-occurrence in bibliographical data extracted from scientific articles indexing in Web of Science, research area "Library & Information Science." We find most frequently software and tools used in bibliometrics have four levels for data treatment: recovery, preparation, processing and analysis, visualization; and use different methods analysis: networks social, geospatial, thematic, temporal, bibliographic coupling, among others.

KEYWORDS

Bibliometrics; Computer programs; Data analysis; Data visualization; Co-word analysis.

Análisis y visualización de datos en estudios bibliométricos

RESUMEN

En los años recientes, académicos de formaciones distintas a las Ciencias de la Información con mayor frecuencia realizan estudios utilizando enfoques y técnicas basados en bibliometría; sin embargo, estos profesionales no siempre tienen los conocimientos y la formación necesarias para realizar estudios métricos de la información de manera rigurosa y metódica. El estudio pretende describir los métodos y herramientas utilizados por investigadores a partir de un análisis de contenido textual empleando métodos de análisis de co-ocurrencia de términos de artículos científicos indizados en *Web of Science*, área de investigación "Bibliotecología". Se encontró que los programas y herramientas informáticas más usadas presentan cuatro niveles de manipulación de datos: recuperación, preparación, procesamiento y análisis, visualización; y emplean distintos métodos de análisis: sociogramas, geoespacial, temático, temporal, acoplamiento, entre otros.

PALABRAS CLAVE

Bibliometría; Programas informáticos; Análisis de datos; Visualización de datos; Análisis de Co-ocurrencia de términos.

1. Introducción

El uso de tecnologías emergentes para el consumo masivo de información y el procesamiento de grandes cantidades de datos se ha incrementado en todas las áreas del conocimiento; resulta imperioso conocer y utilizar herramientas y programas informáticos para fines de investigación y enseñanza en el camino hacia la construcción de nuevo conocimiento. En el ámbito académico, estudiantes universitarios, docentes y académicos emplean herramientas tecnológicas para recuperar, procesar y analizar la información, a fin de comunicar y representar de forma objetiva los resultados de sus hallazgos y conocimientos. En este contexto, observamos que gestores de investigación y consultores independientes utilizan enfoques y técnicas basados en estudios métricos de la información (i.e. bibliometría) no necesariamente para fines académicos; sin embargo, estos profesionales no siempre tienen la formación necesaria o las capacidades específicas para utilizar las herramientas bibliométricas de modo adecuado (Vílchez-Román y Alhuay-Quispe 2016).

En la última década, el crecimiento de estudios en Bibliotecología y Ciencias de la Información muestra un estado progresivo, en específico, en estudios métricos de la información como la bibliometría (Figura 1). Destacan la diversificación en el uso de enfoques y fundamentos teóricos, métodos y modelos de análisis, así como herramientas informáticas para la representación de los datos resultantes del análisis bibliográfico; no obstante, pocos se han detenido a estudiar las características de los programas informáticos que permiten una adecuada presentación de datos.

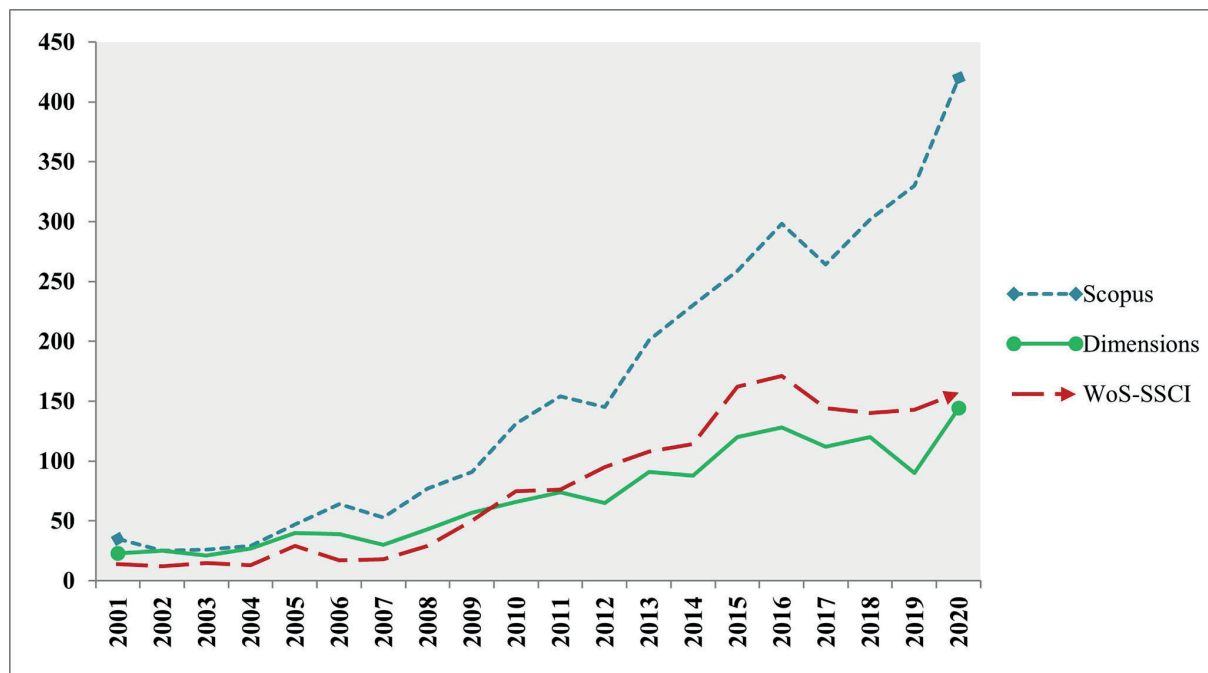


Figura 1. Evolución de producción de artículos sobre “bibliometría” en Scopus, Web of Science y Dimensions.

2. Orígenes y nexos entre disciplinas métricas de información

Las disciplinas métricas de la información se han diversificado y adquirido un *corpus* teórico y metodológico especializado. El antecedente directo de la bibliometría, por ejemplo, fue el término *bibliografía estadística* usada en 1922 por E. Wyndham Hulme en una conferencia titulada *Standards Reader in Bibliography* presentada en la Universidad de Cambridge (Pritchard 2003); pero no sería sino hasta 1944 que el término volvió a ser usado por Gosneell en su artículo *Obsolescence of books in college libraries* (Gosnell 1978). Pritchard (2003) recomienda el uso de “bibliometría” para referirse a la aplicación de métodos estadísticos a los libros y a otros medios de comunicación; mientras que Tague-Sutcliffe (1992) indica que la bibliometría engloba el estudio de los aspectos cuantitativos de la producción, disseminación y uso de la información registrada para elaborar modelos matemáticos para realizar pronósticos y posteriormente tomar decisiones.

La Cienciometría o *Naukometriya*, por su origen ruso, tiene como precursores a Dobrovoy y Nalimov en 1966 y 1969, respectivamente (Callon, Courtial, y Penan 1995); aunque otro posible origen se asocia al surgimiento de la revista *Scientometrics* en 1978, que marcó un hito importante en el ámbito de las entonces conocidas como ciencias cuantitativas. Spinak (1998, 142) sostiene que la cienciometría aplica técnicas bibliométricas a la ciencia y examina el desarrollo y las políticas científicas para establecer comparaciones entre las políticas de investigación entre los países analizando sus aspectos económicos y sociales.

Tague-Sutcliffe (1992) atribuye el surgimiento del término Informetría, al alemán Otto Nacke, quién lo habría propuesto por primera vez en el año 1979, y afirma además que la informetría abarca el estudio de los aspectos cuantitativos de la información, independientemente de la forma en que aparezca registrada y del modo en que se genere y puede incorporar y utilizar diversos medios en la medición de la información, que están fuera de los límites de la bibliometría y de la cienciometría.

El término “webmetría” aparece originalmente como *Webometrics* en inglés cuyo término fue acuñado por Almind y Ingwersen (1997). La webmetría consiste en la aplicación de los métodos informétricos a la *World Wide Web* y puesto que la web se está tornando en un medio importante para la comunicación de la ciencia, es lógico que los estudios cuantitativos se orienten a este ámbito (Vanti 2002). Una de las aplicaciones reales de los métodos basados en elementos web se pueden encontrar en el Ranking Web de Universidades (Webometrics.info), el cual en el ámbito latinoamericano posee gran aceptación y relevancia, ya que no solo incluye indicadores de medición web como presencia y visibilidad web sino también recoge indicadores bibliométricos utilizados en la web académica.

Recientemente *Altmetrics* o métricas alternativas aparecen como un estudio de nuevas formas de medición del alcance e impacto de las publicaciones científicas en el entorno de la social media. Altmetrics abarca el estudio de la actividad científica a partir de nuevos indicadores basados en fuentes de información de la web social, considerando el uso (*views, downloads*), interacciones y recomendaciones (comentarios, favoritos), publicaciones y compartidos (*posts, re-tuits* y *microblogging*), almacenamiento de lecturas (*bookmarking*) y hasta los comentarios de evaluación pública por pares (Uribe-Tirado y Alhuay-Quispe 2017). Las métricas alternativas surgen para proporcionar una visión complementaria al análisis de citas, asumiendo un papel en la evaluación de la ciencia (Borrego 2014).

Guzmán Sánchez (1999, 41) es de los pioneros de emplear el término Patentometría en español, quien entiende que abarca dos aspectos: la producción de conocimientos certificados y la participación en la elaboración de procesos industriales, en este último emplea cuatro indicadores de medición: actividad, relacionales de primera generación, relacionales de segunda generación, y familia de patentes. La utilización eficiente de la información contenida en las patentes es la principal herramienta para establecer tendencias en el registro de la propiedad intelectual, innovación o novedades que pueden producir cambios o surgimientos abruptos de tecnologías (Díaz, Giráldez, y Carrillo 2016; Verma y Suri 2021).

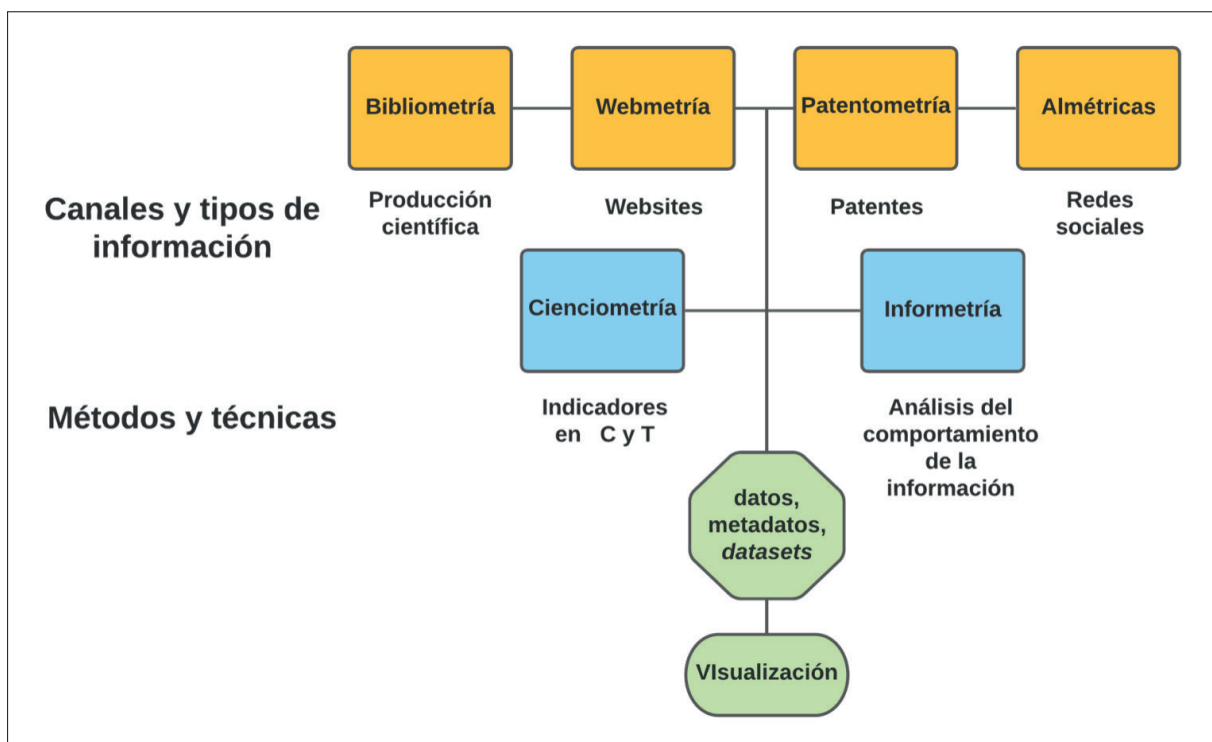


Figura 2. Propiedades de las disciplinas métricas para la visualización

Una breve descripción de los orígenes de cada disciplina métrica nos permite introducirnos de manera panorámica a los elementos que conforman las unidades de estudio, las técnicas de análisis utilizada, los enfoques metodológicos y los fundamentos teóricos, así como, conocer las relaciones y características propias del flujo de la información y comprender la importancia del estudio del tratamiento de los datos en la actividad informacional en cada disciplina métrica. En ese sentido, en la Figura 2 se muestran los elementos clave de cada disciplina métrica, en donde encontramos que algunas se centran en el análisis de canales de información determinados (web, redes sociales), mientras que otras se centran en fuentes de información específicas (artículos científicos, libros) que fueron el principal soporte de estudio en los inicios de la bibliometría. En todos los casos, las disciplinas desarrollaron métodos y técnicas para el análisis de metadatos bibliográficos y conjuntos de datos como insumos de entrada para la visualización de los datos analizados.

3. Visualización de datos e información

La visualización de datos no es una actividad reciente, según Friendly (2008), tiene una historia de muchos siglos atrás; que se remonta desde los primeros mapas y diagramas (antes del siglo XVII), medición y teoría (1600-1699), nuevas formas gráficas (1700-1799), el comienzo de los gráficos modernos (1800-1850), la edad de oro de los gráficos estadísticos (1850-1900), la Edad Media Moderna (1900-1950), Renacimiento de la visualización de datos (1950-1975) hasta la visualización de datos dinámica, interactiva y de alta resolución.; aunque la periodización que establece el autor, según propias palabras, es algo artificial pero brinda la oportunidad de caracterizar los logros en cada período. En este último periodo, Tufte (2016) publica *The Visual Display of Quantitative Information*, un libro popular e innovador sobre la visualización de datos.

La visualización de datos se ha utilizado para ayudar a la comprensión, recopilar datos e información, realizar análisis y comunicar evaluaciones en diversos entornos (Azzam et al. 2013). Actualmente todos los campos de estudio analizan y procesan grandes cantidades de datos (*big data*) que podemos denominar macrodatos donde la visualización de datos juega un rol fundamental para impulsar análisis complejos y demostrar la solidez de las investigaciones a través de presentaciones gráficas.

Por otro lado, la visualización de información surge de distintas áreas con enfoque multidisciplinario como la computación científica (Kim, Zhu, y Chen 2016), psicología (Sánchez-Bonvehí y Ribera 2014) y estadística (Valero Sancho, Català Domínguez, y Marín Ochoa 2014); sin embargo, el concepto abarca el estudio de la transformación de datos, información y conocimiento en representaciones visuales interactivas generadas por computadora (Chen 2010; Liu et al. 2014). Desde una perspectiva actual, la visualización de información es la conformación de una imagen mental a partir de un concepto abstracto y se obtiene a partir de la representación gráfica de variables asociadas al concepto que se quiere visualizar (Guzmán Sánchez, Calero, y Villaseñor 2008), y que emerge para cambiar el proceso comunicativo y devolver estratégicamente el protagonismo a lo visual; así mismo, permite profundizar y aprovechar las ventajas que ofrece la percepción humana y la interacción, con el objetivo de transformar realidades complejas y abstractas en realidades simples (Sánchez-Bonvehí y Ribera 2014).

La visualización de información bibliométrica es una consecuencia del surgimiento de una serie de metodologías y herramientas con fines específicos para un área determinada, una base de datos señalada, indicadores limitados y uso de software para procesamiento y análisis creados con propósitos diferentes al análisis bibliométrico (Sotolongo Aguilar, Guzmán Sánchez y Carrillo 2002). Paralelamente surgen otros enfoques apoyados en el tratamiento y análisis de datos e información, tanto para su uso, comprensión y comunicación en entornos de la academia, universidad y en la industria. La visualización a partir de datos pretende construir un conjunto gráfico y sintético que permita entender, establecer agrupaciones, relaciones o tendencias estadísticas y facilite la obtención de conclusiones para su interpretación (Valero Sancho, Català Domínguez, y Marín Ochoa 2014). El uso de técnicas de visualización de datos se ha diversificado en áreas como cartografía, demografía, ciencias de los datos; sin embargo, su desarrollo destaca en la cartografía en línea, cuya combinación con técnicas como minería de datos, web semántica, datos abiertos enlazados, y otras técnicas avanzadas de visualización, permiten acceder y poner en valor el conocimiento científico (Martín-Forero Morente, Barriuso Mediavilla, y Del-Bosque 2018).

La visualización científica es una forma muy útil para extraer información importante de forma concisa y rápida (Cadahía Subiñas 2016). Se basa en el uso de imágenes y en la transformación de datos o información en imágenes para explicar y comunicar ideas (De La Cruz Salas 2005). Como herramienta para la representación visual de un conjunto de datos que facilita su efectivo análisis y evaluación, permite a los científicos computacionales analizar, entender y comunicar datos numéricos generados durante una investigación (Yatagama Lokuge 2015).

De La Cruz Salas (2005) distingue dos grupos, la visualización informativa y visualización científica – propiamente –, donde la primera hace referencia a aquellos datos que no precisan de una estructuración espacial (e.g. datos que se representen con gráficos de barras), mientras que la visualización científica precisa de un contexto espacial para representarse con claridad: representa elementos reales que precisan un espacio sobre el cual ser visualmente integrado.

El estudio pretende identificar y analizar las características de los programas informáticos empleados para el procesamiento y visualización de datos en estudios bibliométricos de Ciencias de la Información.

4. Métodos

La investigación es de carácter descriptiva y analítica con enfoque cuantitativo y corte transversal. Emplea la técnica de análisis de términos y toma como unidades de análisis a datos bibliográficos textuales (título, resumen, título y palabras clave) de artículos publicados en revistas indizadas en Web of Science, colección principal.

Los análisis de términos implican la extracción de palabras clave a partir de los registros documentales en bases de datos de indización, pero también se pueden extraer las palabras clave de un documento textual, a partir del cual construir una matriz de palabras clave (He 1999). Posteriormente se procede a analizar la matriz de datos textuales (*corpus*) con el propósito de identificar características relevantes de los datos que permitan responder preguntas de investigación a partir de redes de colaboración, términos de tendencia, relaciones y jerarquías de conceptos, polarización de tópicos, entre otros métodos. Se entiende por polarizar los problemas de investigación a que un término puede estar relacionado con muchas otras palabras, por lo que será particularmente importante para la definición de problemas centrales (Courtial, Callon, y Sigogneau 1984).

Se emplea como técnica de recolección de datos a la revisión documental que incluye la consulta de fuentes de tipo secundario: *Web of Science* (WoS), y otros sitios web. La recuperación de datos de WoS se basa en la búsqueda avanzada bajo dos variables agrupadas por el operador “OR” y asociadas por “AND” que se detalla en el Material Suplementario. En la consulta efectuada se emplean límites de tipo documental (*article, review*) y fecha de publicación (2008-2018) pero sin restricción de idioma de la publicación. Se obtuvieron cantidades distintas de acuerdo con las colecciones principales WoS: SCI-Expanded (n=600); SSCI (n=622); ESCI (n=214); A&HCI (n=12). De los resultados se excluyen la colección ESCI por la cobertura reciente (2015-actual) y A&HCI por la reducida cantidad encontrada que asumimos no representativa para el estudio. Los resultados de ambas colecciones restantes (SCI y SSCI) fueron combinadas obteniéndose 848 artículos, siendo esta cantidad asumida como la muestra del estudio, de carácter no probabilístico y tipo intencional o criterial.

Las unidades de análisis (n=848) fueron exportadas en formato de la propia base de datos (*vos tagged format*) en fichero .txt para la preparación del corpus. El proceso de pre-preparación de los datos textuales emplea el análisis de frecuencia de términos usando dos herramientas de análisis textual: TAPoRWare (*Text Analysis Portal for Research* - <http://taporware.ualberta.ca>) y Textanz (<http://www.textanz.com>).

Para revelar la presencia y medir la magnitud (intensidad) de los términos se calcula la frecuencia con la que cada palabra clave está asociada entre sí y se dice que existe una conexión cuando esta frecuencia supera cierta probabilidad de asociación ($p > 0,5$), que se calcula desde la palabra más frecuente hacia la menos frecuente considerando que la palabra más frecuente representa una postura de interés más general, es decir, su poder para polarizar es mayor siendo esta probabilidad llamada coeficiente de inclusión (Courtial, Callon, y Sigogneau 1984).

La co-ocurrencia permitió identificar grupos de términos resultantes de cada programa, los cuales fueron revisados manualmente del más frecuente al menos recurrente con el fin de reconocer nombres de programas o aplicaciones web que hayan sido empleados en los artículos bibliométricos analizados. Luego se seleccionaron los disponibles con sitios web en Internet y cuenten con estas características:

- posibilidad de descarga (prueba o gratuito),
- información de guía de usuario,
- documentación publicada como artículo, libro u otro.

Las variantes de nombres del conjunto de programas y aplicativos informáticos identificados de la etapa de pre-procesamiento fueron normalizadas usando un tesoro de términos que fue construido para tal propósito. Finalmente, se realizaron análisis posteriores de frecuencia y correlación de términos sobre la base del corpus de texto y el listado de nombres de programas, empleando el programa VantagePoint.

5. Resultados

6. Análisis descriptivo: programas y frecuencias de aparición

Se encontraron veinte programas de computadora y aplicativos en línea empleados como herramientas en estudios bibliométricos, algunos desarrollados por programadores o investigadores particulares dedicados al estudio de la ciencia; y otros, desarrollados por instituciones vinculadas a los mismos ámbitos. En la Tabla 1 se muestran detalles principales de los programas como desarrolladores del software y la página de descarga respectiva.

Tabla 1. Lista de programas informáticos utilizados en estudios bibliométricos

N°	Nombre del programa	Desarrollador	Enlace de descarga
1	BibExcel	Olle Persson, Universidad de Umea – Suecia	http://homepage.univie.ac.at/juan.gorraiz/bibexcel/index.html
2	CiteSpace II	Universidad de Drexel – USA	http://cluster.ischool.drexel.edu/~cchen/citespace/download/
3	CitNetExplorer	Universidad de Leiden - Países Bajos	http://www.citnetexplorer.nl/Download
4	CoPalRed	Rafael Bailón-Moreno. Universidad de Granada	http://ec3.ugr.es/copalred/
5	Egonet	Christopher McCarthy. Universidad de la Florida – USA	http://www.arschile.cl/egonet/
6	HistCite	Eugene Garfield	http://interest.science.thomsonreuters.com/forms/HistCite/
7	Loet Leydessdorf	Universidad Amsterdam	http://www.leydesdorff.net/
8	Meta Knowledge	John McLevey y Reid McIlroy-Young	http://networkslab.org/metaknowledge/
9	Network Workbench	Universidad de Indiana	http://nwb.cns.iu.edu/download.html
10	Pajek	Vlado	http://vlado.fmf.uni-lj.si/pub/networks/pajek/
11	Publish or Perish	Tarma Software Research	http://www.harzing.com/download/
12	Scholarometer	Indiana University	http://scholarometer.indiana.edu/
13	Science of Science	Universidad de Indiana	https://sci2.cns.iu.edu/user/index.php
14	Ucinet NetDraw	Lin Freeman, Martin Everett y Steve Borgatti	https://sites.google.com/site/ucinetsoftware/
15	VosViewer	Universidad de Leiden - Países Bajos	http://www.vosviewer.com/Home
16	SciMAT	University of Granada. Soft Computing and Intelligent Information Systems	https://sci2s.ugr.es/scimat/
17	Gephi	Mathieu Bastian	https://gephi.org/
18	CitedReferencesExplorer	Andreas Thor, University of Applied Sciences for Telecommunications Leipzig	http://www.crexplorer.net/
19	STICCI	Valentín Gómez-Jáuregui, Cecilia Gómez-Jáuregui	http://www.sticci.eu/
20	Sitkis	Henri Schildt, Helsinki University of Technology	https://sites.google.com/site/sitkisbibliometricanalysis/home

Fuente: Elaboración propia.

La Tabla 2 muestra un análisis descriptivo de frecuencia de aparición de los nombres de los programas, a partir de la base de datos de registros bibliográficos recuperados de la base de citación, el cual arrojó mayor predominancia (50.4%) de presencia en estudios bibliométricos a tres programas: *VosViewer*, *CiteSpace II* y *Ucinet NetDraw*. El análisis de co-ocurrencia de términos independientes confirma la preferencia de uso para dos primeros programas observados con el conteo por registros, pero aparecen *Pajek*, *HistCite* y *Publish or Perish* entre los primeros siguientes con 10% de frecuencia cada uno (Tabla 2).

Tabla 2. Frecuencia de aparición de los términos de programas usados

Nombre del programa	Número de registros				Número de ocurrencias			
	WoS	%	VP	%	TT	%	TW	%
VosViewer	51	26,02	64	35,56	80	40,82	80	40,82
CiteSpace II	26	13,27	33	18,33	48	24,49	47	23,98
Ucinet NetDraw	22	11,22	14	7,78	21	10,71	21	10,71
Pajek	20	10,20	14	7,78	20	10,20	21	10,71
HistCite	19	9,69	14	7,78	27	13,78	1	0,51
Publish or Perish	17	8,67	6	3,33	21	10,71	20	10,20
BibExcel	12	6,12	8	4,44	12	6,12	13	6,63
SciMAT	7	3,57	4	2,22	8	4,08	9	4,59
Loet Leydesdorff	5	2,55	2	1,11	2	1,02	3	1,53
Gephi	3	1,53	7	3,89	6	3,06	5	2,55
Science of Science	3	1,53	1	0,56	2	1,02	1	0,51
CitNetExplorer	2	1,02	5	2,78	8	4,08	7	3,57
Sitkis	2	1,02	1	0,56	1	0,51	1	0,51
CoPalRed	1	0,51	1	0,56	1	0,51	1	0,51
Egonet	1	0,51	1	0,56	4	2,04	4	2,04
Meta Knowledge	1	0,51	1	0,56	4	2,04	3	1,53
Scholarometer	1	0,51	0	0,00	1	0,51	1	0,51
STICCI	1	0,51	1	0,56	2	1,02	2	1,02
CitedReferencesExplorer	1	0,51	2	1,11	3	1,53	2	1,02
Network Workbench	1	0,51	1	0,56	1	0,51	1	0,51

VP: VantagePoint. TT: TexTanz. TW: TaporWare.

7. Nivel de manipulación y análisis de datos

En los estudios métricos exhaustivos y metódicos, el proceso que conduce a la visualización de datos analizados tiene cuatro niveles predecesores: a) recuperación de los datos a analizar; b) preparación de los datos recolectados, esto implica llevarlos a formatos compatibles con otros programas; c) procesamiento y análisis consiste en el empleo de algoritmos, métodos y técnicas para su posterior representación de los datos y d) la visualización, la parte final del proceso. De estas cuatro, en la Tabla 3 se muestra que la etapa visualización posee la mayor frecuencia (n=15); y son pocos (n=3) los que abarcan la mayor cantidad de etapas, de las cuales figuran: Science of Science, Meta Knowledge, SciMAT. De igual forma, tres aplicativos permiten la recuperación de información, siendo PoP el que no abarca ninguno de los demás niveles de manipulación de datos.

Tabla 3. Nivel de manipulación de los datos

N°	Nombre del programa	REC	PRE	PRO	VIS
1	BibExcel		x		
2	CiteSpace II				x
3	CoPalRed			x	
4	Loet Leydessdorf		x		
5	Network Workbench			x	x
6	Science of Science		x	x	x
7	VosViewer				x
8	HistCite				x
9	CitNetExplorer				x
10	Pajek			x	x
11	Pop	x			
12	Egonet			x	x
13	Scholarometer	x			x
14	Ucinet NetDraw			x	x
15	Meta Knowledge	x		x	x
16-	SciMAT		x	x	x
17-	Gephi			x	x
18-	Cited References Explorer				x
19	STICCI		x	x	
20	Sitkis			x	x

REC: Recuperación, PRE: Preparación, PRO: Procesamiento y análisis, VIS: Visualización

Se encontró que los programas son capaces de emplear diversos métodos como el análisis de redes, análisis geoespacial, análisis temático, análisis temporal, acoplamiento bibliográfico, co-ocurrencia, etc. En la Tabla 4 se muestra un análisis de correlación de términos agrupados como conceptos, un *ranking* de diez conceptos identificados donde se destacan los análisis de citación y de impacto científico.

Tabla 4. Matriz de correlación de términos por métodos de análisis de los datos

k	Términos	1	2	3	4	5	6	7	8	9	10
121	1. Citation Analysis	1,00	0,25	0,27	0,25	0,08	0,10	0,00	0,13	0,06	0,00
119	2. Scientific Impact		1,00	0,21	0,24	0,08	0,00	0,00	0,07	0,00	0,00
106	3. Bibliometric Indicators			1,00	0,32	0,08	0,00	0,00	0,25	0,10	0,00
80	4. Hirsch-Index				1,00	0,06	0,00	0,00	0,13	0,00	0,00
59	5. Social Network Analysis					1,00	0,30	0,20	0,05	0,16	0,10
33	6. Co-Citation Analysis						1,00	0,13	0,00	0,07	0,17
27	7. Co-Word Analysis							1,00	0,00	0,00	0,00
25	8. Research Performance								1,00	0,00	0,00
23	9. Collaboration									1,00	0,00
17	10. Data Visualization										1,00

K=número de ocurrencias.

La evaluación realizada encontró que las fuentes de datos empleadas indistintamente son: Web of Science, JCR, Scopus, PubMed, arXiv, Google Scholar, NASA ADS y National Science Foundation. La Tabla 5 muestra que Web of Science es la fuente de datos más recurrente, seguido de Scopus, lo que evidencia que actualmente son la autoridad en información científica internacional; y en menor proporción: PubMed y Google Scholar.

Tabla 5. Matriz de correlación de términos por fuente de entrada de datos

k	Términos	1	2	3	4	5
32	1. WoS	1,00	0,30	0,25	0,00	0,00
29	2. Scopus		1,00	0,26	0,00	0,11
18	3. Google Scholar			1,00	0,12	0,27
4	4. PubMed				1,00	0,00
3	5. GSM					1,00

K=número de ocurrencias. GSM=Google Scholar Metrics.

8. Discusión y conclusiones

La literatura reporta que diversos estudios proponen instrumentos y metodologías para el análisis de programas informáticos, otros estudian las metodologías empleadas en estudios métricos; sin embargo, los pormenores del tratamiento de los datos en investigaciones bibliométricas o de otras disciplinas métricas son escasas. En el primer enfoque, (Macías Rivero, Guzmán Sánchez, y Martínez Suárez 2009) diseñan un modelo de evaluación de software que aplica indicadores métricos para el ejercicio de la vigilancia científico-tecnológica, tomando como guía la ISO: 9126-1 (2001), y proponen indicadores que explican cómo tabular y describir cada software. En el segundo enfoque (Guzmán Sánchez, Calero, y Villaseñor 2008) se evalúan diferentes interfaces visuales a partir de un modelo de comparación, a fin de obtener información comparativa de las fases de entrada de los datos, análisis y proyección, para comprender mejor el resultado total de la visualización. En un tercer enfoque (Cobo et al. 2011) se analizan y comparan técnicas bibliométricas y programas empleados para el mapeo de la ciencia. Asimismo, Alfonso, Sakraida, y Hastings-Tolsma (2014) emplean el mapeo bibliométrico como método de representación y visualización de datos bibliométricos; además, revisan las consideraciones metodológicas de los programas de computadora empleados para el análisis de citas.

La representación o visualización de los datos resultaría una actividad casi imposible, más aún si se pretende analizar íntegramente la totalidad de los datos, comparar temporalmente el comportamiento e interrelación de los mismos o simplemente servir de instrumento didáctico de transmisión de conocimiento. Aunque el proceso de visualización y representación de datos tiene su origen en la presentación gráfica de datos en la Estadística, en la actualidad, la interacción entre la Matemática, Estadística y Ciencias de la Información se ha visto potenciado con el desarrollo de las Ciencias de la Computación y otros métodos, procesos y tecnologías particulares de éstas, relacionados con el tratamiento, almacenamiento y análisis de los datos (Gorbea Portal 2013), por lo que en el campo del estudio de la ciencia, el uso de métodos de análisis informétricos constituyen

una herramienta indispensable; sin embargo, la presentación tradicional de resultados de datos procesados en el análisis de citas, co-citaciones, co-autoría, redes sociales, etc., a manera de tablas planas, gráfico de barras, diagramas polares, entre otros, resulta bastante limitante al momento de representar datos analizados en un estudio bibliométrico y cuantitativo.

La representación de datos en estudios métricos permite su comprensión didáctica, y la representación visual de grandes cantidades de datos. Sin embargo, la exponencial creciente y explosión de la producción científica, materia de estudio del análisis métrico de la información, así como, administrar una gran cantidad de datos, desde su recuperación a través de diversas bases de datos bibliográficas, como su procesamiento y análisis, a partir de indicadores, variables y modelos bibliométricos, resultaría una actividad casi épica; además del riesgo en la veracidad de los resultados por la posible omisión o pérdida de datos que pudiera surgir en el proceso manual de la recuperación y análisis.

La visualización de datos como proceso último de la cadena de análisis de datos emplea herramientas informáticas de acuerdo con el propósito del estudio en particular. Estos estudios pueden orientarse al *análisis de redes sociales* (Molina, Muñoz, y Domenech 2002; Delgado López-Cózar et al. 2006; Repiso, Torres-Salinas, y Delgado López-Cózar 2011a; 2011b), *mapeo historiográfico* (Garfield 2004), *mapas bibliométricos* (Eck y Waltman 2010; 2007; Van Raan y Van der Velde 1991), entre otros. Algunos estudios previos (Cobo et al. 2011; Alfonzo, Sakraida, y Hastings-Tolsma 2014) demuestran que los programas informáticos empleados para el análisis y visualización de datos de estudios métricos son diversos, así como los métodos y técnicas de procesamiento, algunos más complejos que otros; de igual modo con las técnicas y algoritmos empleados por los programas para la representación gráfica de datos.

En conclusión, los programas informáticos empleados ya sea para el procesamiento o para la visualización de los datos bibliográficos en los artículos de análisis bibliométricos de Ciencias de la Información estudiados presentan hasta cuatro niveles de manipulación de los datos de entrada: i) recuperación, ii) preparación, iii) procesamiento y análisis, iv) visualización, en los que se emplean diferentes métodos y técnicas como los análisis de sociogramas, co-ocurrencias y acoplamiento bibliográfico, entre los más recurrentes. Respecto a los datos de entrada, estos provienen de varias fuentes externas de información bibliográfica y documental (i.e. bases de datos científicas, repositorios y buscadores), que en algunos casos los programas incluyen como funcionalidad, mientras que otros permiten la extracción masiva de datos de fuentes secundarias (Web of Science, JCR, Scopus, PubMed, arXiv, Google Scholar. Esta investigación abordó un aspecto que se refiere al proceso inicial de recuperación de datos; sin embargo, resulta importante el diseño de un modelo de evaluación de programas informáticos que vincule las características del manejo de datos en estudios métricos; además de modelos de análisis de interfaces visuales (Macías Rivero, Guzmán Sánchez, y Martínez Suárez 2009; Guzmán Sánchez, Calero, y Villaseñor 2008). Estudios futuros podrían indagar sobre la naturaleza y características del tratamiento y visualización de datos en las demás disciplinas métricas de la información de modo que la evidencia encontrada permita conocer y superar las limitaciones metodológicas frente al crecimiento masivo de los datos en el contexto de *big data* y facilite la utilización de los recursos e insumos de la producción científico e información documental para una adecuada toma de decisiones en la gestión de la información, de la investigación y del conocimiento explícito en las instituciones de educación superior y organizaciones corporativas.

Referencias bibliográficas

- Alfonzo, Paige M., Teresa J. Sakraida, y Marie Hastings-Tolsma. 2014. «Bibliometrics: Visualizing the Impact of Nursing Research». *Online Journal of Nursing Informatics* 18 (1). <http://ojni.org/issues/?p=3093>.
- Almind, Tomas C., y Peter Ingwersen. 1997. «Informetric Analyses on the World Wide Web: Methodological Approaches to “Webometrics”». *Journal of Documentation* 53 (4): 404-26. <https://doi.org/10.1108/EUM0000000007205>.
- Azzam, Tarek, Stephanie Evergreen, Amy A. Germuth, y Susan J. Kistler. 2013. «Data Visualization and Evaluation». *New Directions for Evaluation* 2013 (139): 7-32. <https://doi.org/10.1002/ev.20065>.
- Borrego, Angel. 2014. «Altmétricas para la evaluación de la investigación y el análisis de necesidades de información». *Profesional de la Información* 23 (4): 352-58. <https://doi.org/10.3145/epi.2014.jul.02>.
- Cadahía Subiñas, Alberto. 2016. «Aplicación web enriquecida para visualización científica». Madrid, España: Universidad Politécnica de Madrid. <http://oa.upm.es/43533/>.
- Callon, Michel, Jean-Pierre Courtial, y Hervé Penan. 1995. *Cienciometría: la medición de la actividad científica: de la bibliometría a la vigilancia tecnológica*. Gijón, España: Ediciones Trea.
- Chen, Chaomei. 2010. «Information Visualization». *Wiley Interdisciplinary Reviews: Computational Statistics* 2 (4): 387-403. <https://doi.org/10.1002/wics.89>.
- Cobo, Manuel J., Antonio G. López-Herrera, Enrique Herrera-Viedma, y Francisco Herrera. 2011. «Science Mapping Software Tools: Review, Analysis, and Cooperative Study among Tools». *Journal of the American Society for Information Science and Technology* 62 (7): 1382-1402. <https://doi.org/10.1002/asi.21525>.
- Courtial, Jean-Pierre, Michel Callon, y M Sigogneau. 1984. «Is Indexing Trustworthy? Classification of Articles through Co-Word Analysis». *Journal of Information Science* 9 (2): 47-56. <https://doi.org/10.1177/016555158400900201>.
- Delgado López-Cózar, Emilio, Daniel Torres-Salinas, Evaristo Jiménez-Contreras, y Rafael Ruiz-Pérez. 2006. «Análisis bibliométrico y de redes sociales aplicado a las tesis bibliométricas defendidas en España (1976-2002): temas, escuelas científicas y redes académicas». *Revista Española de Documentación Científica* 29 (4 SE-Estudios): 493-524. <https://doi.org/10.3989/redc.2006.v29.i4.306>.
- Díaz, Maidelyn, Raudel Giráldez, y Humberto Carrillo. 2016. «La Patentometría como herramienta de vigilancia para monitorear las colaboraciones tecnológicas de un dominio». En *III Congreso Internacional de Información – INFO´ 2016, La Habana, Cuba*. La Habana, Cuba. <http://www.congreso-info.cu/index.php/info/2016/paper/view/44/50>.
- Eck, Nees Jan van, y Ludo Waltman. 2007. «Bibliometric Mapping of the Computational Intelligence Field». *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 15 (5): 625-45. <https://doi.org/10.1142/S0218488507004911>.

- . 2010. «Software Survey: VOSviewer, a Computer Program for Bibliometric Mapping». *Scientometrics* 84 (2): 523-38. <https://doi.org/10.1007/s11192-009-0146-3>.
- Friendly, Michael. 2008. «A Brief History of Data Visualization BT - Handbook of Data Visualization». En *Handbook of Data Visualization*, editado por Chun-houh Chen, Wolfgang Härdle, y Antony Unwin, 15-56. Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-33037-0_2.
- Garfield, Eugene. 2004. «Historiographic Mapping of Knowledge Domains Literature». *Journal of Information Science* 30 (2): 119-45. <https://doi.org/10.1177/0165551504042802>.
- Goarbea Portal, Salvador. 1994. «Principios teóricos y metodológicos de los estudios métricos de la información». *Investigación Bibliotecológica: archivonomía, bibliotecología e información* 9 (17): 23-32. <https://doi.org/10.22201/iibi.0187358xp.1994.17.3826>.
- . 2013. «Tendencias transdisciplinarias en los estudios métricos de la información y su relación con la gestión de la información y del conocimiento». *Perspectivas em Gestão & Conhecimento* 3 (1): 13-27. <https://periodicos.ufpb.br/ojs2/index.php/pgc/article/view/14175>.
- Gosnell, Charles F. 1978. «Obsolescence of Books in College Libraries». *Collection Management* 2 (2): 167-82. https://doi.org/10.1300/J105v02n02_07.
- Guzmán Sánchez, María Victoria. 1999. «Patentometría: herramienta para el análisis de oportunidades tecnológicas». Universidad de La Habana. <http://www.bvv.sld.cu/docs/documentos/119566413228.pdf>.
- Guzmán Sánchez, María Victoria, R. Calero, y E. Villaseñor. 2008. «Evaluación de interfaces visuales para estudios métricos: modelo de comparación». En *IV Seminario Internacional sobre Estudios Cuantitativos y Cualitativos de la Ciencia y la Tecnología*. La Habana, Cuba: Instituto de Información Científica y Tecnológica.
- He, Quin. 1999. «Knowledge Discovery Through Co-Word Analysis». *Library Trends* 48 (1): 133. <http://hdl.handle.net/2142/8267>.
- Kim, Meen Chul, Yongjun Zhu, y Chaomei Chen. 2016. «How Are They Different? A Quantitative Domain Comparison of Information Visualization and Data Visualization (2000–2014)». *Scientometrics* 107 (1): 123-65. <https://doi.org/10.1007/s11192-015-1830-0>.
- La Cruz Salas, Luis Miguel De. 2005. «Visualización científica». *Revista Digital Universitaria* 6 (12): 2-9. <http://www.revista.unam.mx/vol.6/num12/art117/int117.htm>.
- Liu, Shixia, Weiwei Cui, Yingcai Wu, y Mengchen Liu. 2014. «A Survey on Information Visualization: Recent Advances and Challenges». *Visual Computer* 30 (12). <https://doi.org/10.1007/s00371-013-0892-3>.
- Macías Rivero, Yaidelyn, María Victoria Guzmán Sánchez, y Yamila Martínez Suárez. 2009. «Modelo de evaluación para software que emplean indicadores métricos en la vigilancia científico-tecnológica». *ACIMED* 20 (6): 125-40. http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1024-94352009001200003.
- Martín-Forero Morente, Lourdes, Alejandro Barriuso Mediavilla, e Isabel Del-Bosque. 2018.

«Webmapping y visualización de datos científicos en las humanidades digitales». En *XVIII Congreso Nacional de Tecnologías de la Información Geográfica: Perspectivas Multidisciplinares en la Sociedad del Conocimiento*, 812-20. Valencia, España: Universidad de Barcelona. <https://digital.csic.es/handle/10261/166984>.

Molina, José Luis, Juan Manuel Muñoz, y Miquel Domenech. 2002. «Redes de publicaciones científicas: un análisis de la estructura de coautorías.» *Redes. Revista Hispana para el análisis de redes sociales* 1 (1). http://ddd.uab.cat/pub/redes/15790185v1/vol1_3.htm.

Pritchard, Alan. 2003. «Bibliografía estadística o bibliometría.» *Infobib*, n.º 2: 33-36.

Raan, Anthony F. J. Van, y J. G. M. Van der Velde. 1991. «Bibliometric Mapping and Early Warning for Emerging Topics: Exploring Developments in Materials Science.» *Research Evaluation* 1 (2): 115-27. <https://doi.org/10.1093/rev/1.2.115>.

Repiso, Rafael, Daniel Torres-Salinas, y Emilio Delgado López-Cózar. 2011a. «Bibliometric and Social Network Analysis Applied to Television: Dissertations Presented in Spain (1976/2007)». *Comunicar* 19 (37): 151-59. <https://doi.org/10.3916/C37-2011-03-07>.

———. 2011b. «Análisis de la investigación sobre Radio en España: una aproximación a través del Análisis Bibliométrico y de Redes Sociales de las tesis doctorales defendidas en España entre 1976-2008.» *Estudios sobre el Mensaje Periodístico* 17 (2): 417-29. https://doi.org/10.5209/rev_ESMP.2011.v17.n2.38123.

Sánchez-Bonvehí, Clàudia, y Mireia Ribera. 2014. «Visualización de la información en la democratización de los datos: propuestas desde el periodismo y la narratividad.» *Profesional de la Información* 23 (3): 311-18. <https://doi.org/10.3145/epi.2014.may.11>.

Sotolongo Aguilar, Gilberto, M^a Victoria Guzmán Sánchez, y Humberto Carrillo. 2002. «ViBlio-SOM: Visualización de información bibliométrica mediante el mapeo autoorganizado.» *Revista Española de Documentación Científica* 25 (4): 477-84. <https://doi.org/10.3989/redc.2002.v25.i4.281>.

Spinak, Ernesto. 1998. «Indicadores cienciométricos.» *Ciência da Informação* 27 (2): 141-48. <http://revista.ibict.br/ciinf/article/view/795>.

Tague-Sutcliffe, Jean. 1992. «An Introduction to Informetrics.» *Information Processing and Management* 28 (1): 1-3. [https://doi.org/10.1016/0306-4573\(92\)90087-G](https://doi.org/10.1016/0306-4573(92)90087-G).

Tufte, Edward. 2016. «The Visual Display of Quantitative Information (1983)». En *The Diagrammatic Reader: Fundamental Theoretical and Historical Texts*, editado por Birgit Schneider, Christoph Ernst, y Jan Wöpking, 219-30. Berlin: De Gruyter. <https://doi.org/10.1515/9783050093833-035>.

Uribe-Tirado, Alejandro, y Joel Alhuay-Quispe. 2017. «Estudio métrico de ALFIN en Iberoamérica: De la bibliometría a las altmetrics.» *Revista Española de Documentación Científica* 40 (3). <https://doi.org/10.3989/redc.2017.3.1414>.

Valero Sancho, José Luis, Jordi Català Domínguez, y Beatriz-Elena Marín Ochoa. 2014. «An Approach to the Taxonomy of Data Visualization.» *Revista Latina de Comunicación Social* 69: 486-507. <https://doi.org/10.4185/RLCS-2014-1021>.

Vanti, Nadia Aurora Peres. 2002. «Da bibliometria à webometria: uma exploração conceitual dos mecanismos utilizados para medir o registro da informação e a difusão do conhecimento». *Ciência da informação* 31 (2): 369-79. <https://doi.org/10.1590/S0100-19652002000200016>.

Verma, Charu, y Pradeep Kumar Suri. 2021. «‘Big data’ patentometrics for R&D decision-making». *Digital Policy, Regulation and Governance* 23 (4): 317-336. <https://doi.org/10.1108/DPRG-09-2020-0126>.

Vílchez-Román, Carlos, y Joel Alhuay-Quispe. 2016. «Use of text mining for understanding Peruvian students and faculties’ perceptions on bibliometrics training». En *CEUR Workshop Proceedings*, 1743:165-69. <http://ceur-ws.org/Vol-1743/paper20.pdf>.

Yatagama Lokuge, y Shamith Wimukthi. 2015. «Extensión del módulo de visualización científica de Weka». Las Villas, Cuba: Universidad Central “Marta Abreu” de Las Villas. <https://dspace.uclv.edu.cu/handle/123456789/6880>.