

# ScatterPE: una herramienta interactiva de apoyo al análisis bibliométrico

José Federico Medrano

Facultad de Ingeniería, Universidad Nacional de Jujuy (UNJu), Jujuy, Argentina.

## Resumen

Las bases de datos bibliográficas poseen una enorme cantidad de registros en forma de publicaciones científico-académicas disponibles y listadas según diversos criterios. Al realizar una búsqueda por autor, estas ofrecen un conjunto de resultados, el cálculo de indicadores y métricas, y en algunos casos opciones de visualización de resultados. Este último aspecto, el que se refiere a la visualización, se encuentra aún en proceso de maduración. Estas herramientas se limitan a presentar simples estadísticas o datos tabulados, dejando de lado algún otro tipo de visualización más compleja que involucre varias dimensiones más que la cantidad de citas recibidas o la cantidad de publicaciones por año. Una adecuada visualización de datos permitiría entender la realidad desde distintas ópticas aportando un mayor entendimiento y un nuevo o mejor conocimiento a veces oculto en representaciones básicas y estáticas. En este trabajo se presenta un prototipo inicial de una herramienta interactiva de visualización de información que servirá de apoyo a los análisis bibliométricos. La idea inicial es intentar mejorar las representaciones visuales actuales contribuyendo con una alternativa de fácil implementación e integración en las aplicaciones actuales.

**Palabras clave:** bibliometría, visualización de información, D3, InfoVis.

## **ScatterPE: an interactive tool to support bibliometric analysis**

### **Abstract**

*Bibliographic databases have a huge number of records in the form of scientific-academic publications available and listed according to various criteria. When performing a search by author these Databases offer a set of results, the calculation of indicators and metrics and, in some cases, options for displaying results. This last aspect, which refers to visualization, is still in the process of maturation. These tools are limited to presenting simple statistics or tabulated data, leaving aside some other type of more complex visualization that involves several dimensions more than the number of citations received or the number of publications per year. Adequate visualization of data would allow understanding the reality from different optics contributing a greater understanding and a new or better knowledge sometimes hidden in basic and static representations. This paper presents an initial prototype of an interactive information visualization tool that will support bibliometric analysis. The initial idea is to try to improve the current visual representations contributing an alternative of easy implementation and integration in the current applications.*

**Keywords:** *bibliometry, information visualization, D3, InfoVis.*

## Introducción

La potencia de las interfaces gráficas y el aumento de la capacidad de procesamiento de los equipos informáticos, permiten entre tantas cosas una interacción directa con la información visualizada, esto ha favorecido la creación y la popularización de métodos y técnicas de Visualización de Información (InfoVis). *InfoVis* busca aumentar el conocimiento humano mediante el aprovechamiento de las capacidades visuales humanas para dar sentido a la información abstracta, **Card et al.** (1999), proporcionando los medios por los cuales los seres humanos mediante sus capacidades perceptivas, pueden lidiar con el constante aumento de la cantidad de datos disponibles.

Las interfaces basadas en texto requieren esfuerzo cognitivo para entender su contenido informativo, por esto la idea básica de la exploración visual de los datos es la de presentar los datos en alguna forma visual, permitiendo que los humanos puedan obtener conocimiento, sacar conclusiones, e interactuar directamente con los mismos. Con este tipo de representaciones basadas en grandes cantidades de datos, los usuarios pueden "detectar patrones o comportamientos que se deseaban evaluar, como así también descubrir comportamientos y relaciones entre los datos desconocidos hasta el momento", **Keim** (2002). *InfoVis* abarca las técnicas de visualización que tienen que ver principalmente con datos abstractos, es decir, los datos para los cuales el usuario no tiene un modelo mental preconcebido. Por esta razón, la interacción es especialmente importante en *InfoVis*, ya sea para la exploración, análisis y/o presentación de los datos, **Kosara et al.** (2003) y **Chen et al.** (2018).

Dado que el sentido primario del ser humano es el sentido de la vista, la mayor parte de la información puede ser transportada utilizando este canal. La interacción permite al usuario implícitamente formar modelos mentales de las correlaciones y las relaciones entre los datos, a través del reconocimiento de patrones, marcando y centrándose en esos patrones, formulando hipótesis y pruebas mentales.

Sin importar el campo de estudio, las visualizaciones, gráficas y representaciones visuales de la información abundan para dar un enfoque distinto a lo ya presentado o simplemente para resumir un conjunto de datos. Dentro del campo de la evaluación de la ciencia, un análisis o estudio bibliométrico parte de un conjunto de registros bibliográficos obtenidos a partir de una

base de datos propia, comercial o de libre acceso, con el objetivo de contabilizar la producción científico-académica de un investigador, grupo de investigación, departamento, institución o país. Los indicadores bibliométricos más empleados y criticados en la evaluación de la investigación son según **Bornmann et al.** (2018), las publicaciones y el recuento de citas.

La *Visualización de Información* cumple un papel relevante al realizar un análisis basado en citas bibliográficas. Poder plasmar mediante una representación, la información recolectada se vuelve una tarea compleja no solo según aumente la cantidad de información, sino también según se incremente el número de dimensiones objeto de estudio.

## Visualizaciones actuales

En la actualidad existen un pequeño número de visualizaciones relacionadas con esta temática, **Liao et al.** (2018); **Yeung** (2018).

*Scopus* (<https://www2.scopus.com>) la base de datos bibliográfica por excelencia, se limita a presentar algunos gráficos de barra, torta y dona para representar los resultados de las búsquedas realizadas, estos gráficos incluyen visualizaciones basadas en la procedencia, tipo, año y área temática de las publicaciones. La única visualización relacionada con el número de citas y con indicadores bibliométricos, es el *h-graph*.

*Knowledge* (<http://apps.webofknowledge.com>) presenta algunas visualizaciones (TreeMap (**Köpp & Weinkauff** (2018)) para analizar los resultados entregados basándose en diferentes agrupaciones como año de publicación, autores, país de publicación, categoría según WoS, entre otras. Además de esto ofrece el informe de citas, la cual es una representación gráfica que muestra la evolución del número de citas recibidas a lo largo de los años.

*Google Scholar* (<https://scholar.google.com>), que es el máximo referente como motor de búsqueda de material científico-académico de libre acceso, **Harzing** (2017), se limita a ofrecer un gráfico de barras de la evolución temporal del número de citas total y de las citas recibidas por cada publicación.

*Microsoft Academic* (<https://academic.microsoft.com>) el gran rival de Google Scholar, ubicado solo un escalón debajo, como alternativa y fuente de datos de

material científico-académico, **Harzing & Alakangas** (2017), se limita a presentar un gráfico de barras donde cada una posee una doble componente, la cantidad de citas recibidas y la cantidad de publicaciones de ese año (observar figura 1).

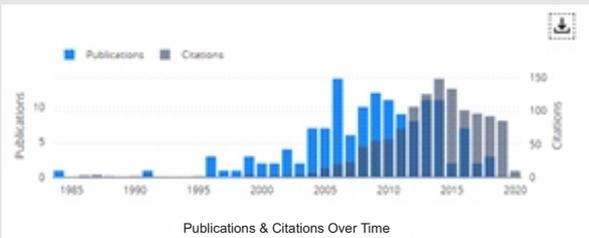


Fig. 1 Número de publicaciones y citas a través del tiempo en Microsoft Academic para un autor.

*Semantic Scholar* (<https://www.semanticscholar.org>), una herramienta de libre acceso lanzada en el año 2015, que fue diseñada como un motor de búsqueda de artículos científicos basado en inteligencia artificial, ofrece un par de gráficos relacionados al número de citas por cada publicación. En el listado de resultados a partir de una búsqueda inicial, ofrece un gráfico de las citas recibidas en los últimos tres años incluyendo el año actual, calculando lo que llama la “velocidad de cita”, que no es otra cosa que el promedio de citas recibidos en estos años.

Otra visualización en esta plataforma es la “Influencia del Autor”, esta es posible ingresando al detalle de publicaciones de un autor, es decir, si la plataforma logró agrupar las publicaciones de un mismo autor, permite ver cuáles son los autores que más influyeron en el autor buscado y cuáles son los autores más influenciados por este (ver figura 2).

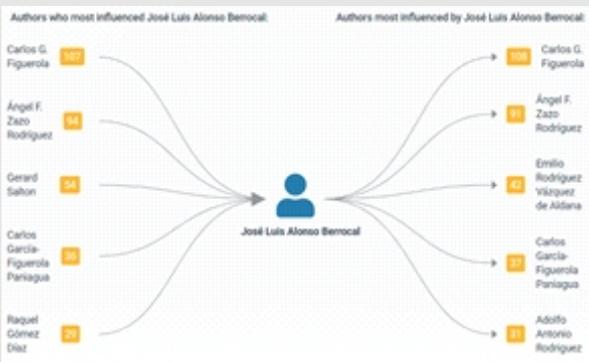


Fig. 2 Gráfico de Influencia de Autor en SemanticScholar.

Un gráfico que intenta reunir las publicaciones de un investigador separándolas por tipo de publicación es el que ofrece el sitio web *CSAuthor*, en la misma se puede observar la evolución a través de los años de la cantidad de publicaciones indicando el tipo de la misma (Libro, En actas de congreso, Artículo, Tesis doctoral, Otros) con un color específico (ver figura 3). Sitio web de *CSAuthor*: <https://www.csauthors.net>.

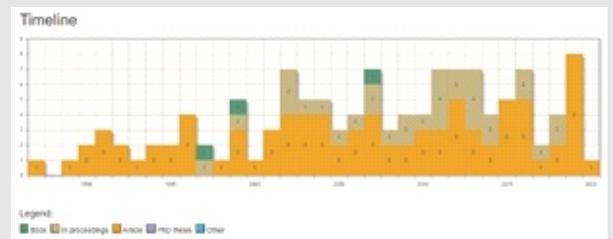


Fig. 3 Evolución temporal del número de publicaciones por tipo de publicación para un autor en *CSAuthor*.

Si bien existen diferentes propuestas para visualizar algunas de las dimensiones que puede llegar a involucrar un análisis bibliométrico, éstas dependen de la herramienta o base de datos bibliográfica utilizada, y mayormente se limitan a representaciones estáticas o en el mejor de los casos solo involucran dos variables, por ejemplo: el año de publicación y la cantidad de documentos publicados por año. Es por esto que en este trabajo se propone el desarrollo de una herramienta interactiva, de libre acceso, versátil, capaz de involucrar un número mayor de variables y ofrecer un conjunto de interacciones en una única visualización, para aportar una mirada desde un prisma diferente los resultados de un estudio bibliométrico.

### Propuesta

El *Scatter plot* también llamado *scatter graph*, *scatter chart*, *scattergram*, *scatter diagram* o *diagrama de dispersión*, es un tipo de diagrama matemático que utiliza coordenadas cartesianas para graficar puntos que muestran la relación entre dos variables de un conjunto de datos. A menudo se utiliza este tipo de diagramas para identificar asociaciones potenciales entre dos variables, en las que se puede considerar una variable explicativa y otra puede considerarse una variable de respuesta, **Lacey** (2017).

Encontrar una visualización que se ajuste a las variables que intervienen en un análisis bibliométrico pue-

de resultar una tarea no tan sencilla, esto se debe a que en este tipo de estudios entran en juego diversas variables, por ejemplo: la cantidad de publicaciones por año, la cantidad de citas recibidas, el tipo de publicación, cantidad de autores, el lugar de publicación, el cálculo de indicadores (h-index y sus variantes). Al respecto, si se decidiera emplear un gráfico de barras donde se evalúan dos variables o hasta una tercera empleando barras apiladas con colores, para un análisis bibliométrico, resultaría en un gráfico muy sobrecargado y hasta difícil de entender según aumente la cantidad de registros. Un gráfico de líneas en un espacio de 2 dimensiones tampoco aporta muchas mejoras a esta problemática, puesto que solo permitiría evaluar dos variables. Con un diagrama de dispersión que permite también evaluar hasta dos variables como se mencionara, la cuestión no cambia, la limitación sigue presente y tampoco sería la mejor opción. Sin embargo, una variante o mejora de este diagrama permite introducir dos variables más, donde una puede estar representada por el tamaño de los puntos graficados, ahora ya no serían puntos sino burbujas, y otra variable puede estar representada por el color de dichas burbujas, en este sentido, se lograría ampliar notablemente las prestaciones del diagrama *Scatter plot* común

al incorporar más dimensiones (variables). Tomando como base este último diagrama, se decidió optar por la variante que involucra cuatro variables de estudio, aunque entrando más en detalle, una quinta variable puede entrar en juego, si bien no como un elemento de visualización sino como un elemento de interacción, que sería a partir de la aplicación de un filtrado por alguna variable a elección. De este modo, un diagrama de dispersión extendido como el que se propone, podría adecuarse de una mejor manera a un mayor número de variables de un estudio bibliométrico.

A partir de lo expuesto, el prototipo construido se basó en un ejemplo sencillo y estático de Scatter plot disponible en el blog de su creador **Bostock** (2019), el cual muestra la relación inversa entre la potencia del motor y la eficiencia del combustible. La idea inicial fue ampliar la funcionalidad de este diagrama en un intento de incorporar la mayor cantidad de variables en una sola visualización y tener así una imagen completa de la situación de estudio, en este caso, un conjunto de publicaciones científico-académicas pertenecientes a un autor determinado recolectadas de una base de datos bibliográfica (el resultado obtenido se observa en la figura 4).

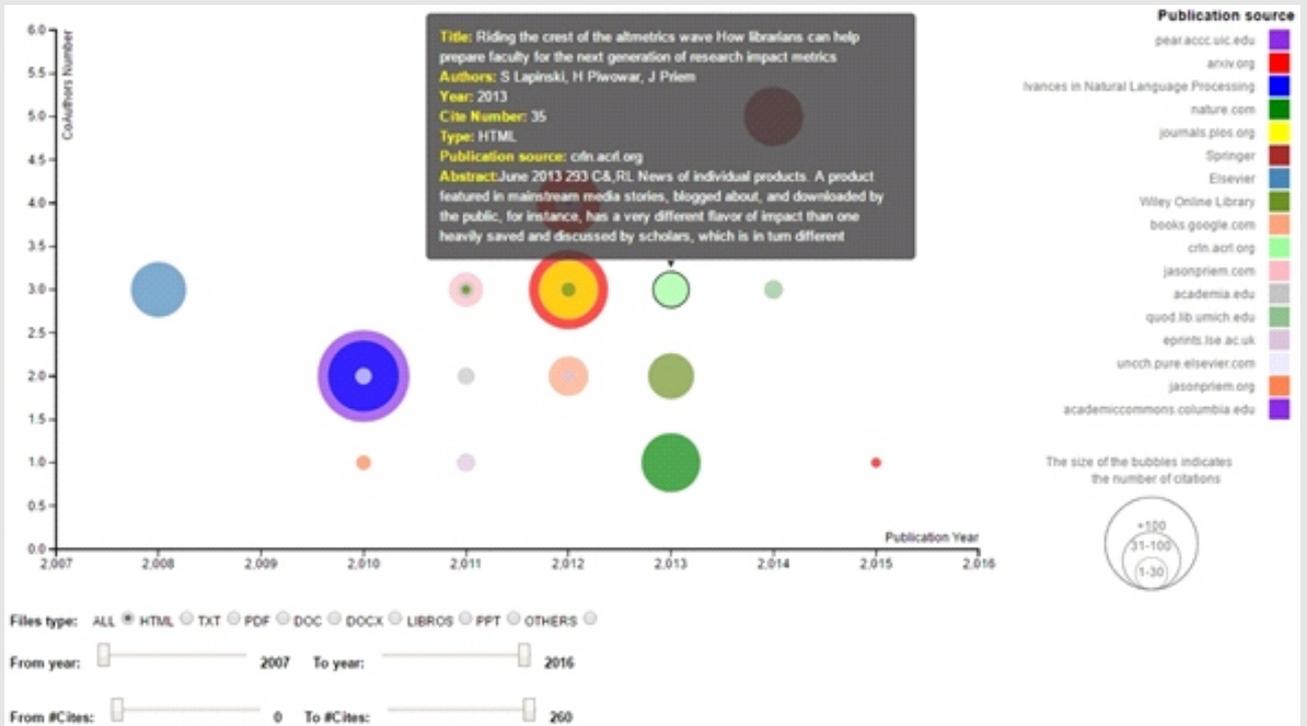


Fig. 4 Visualización de ScatterPE para los registros de un autor tomando como origen de datos Google Scholar.

Para el diseño, modificación e implementación de la visualización presentada, se realizaron diversos experimentos con conjuntos de datos provenientes de motores de búsqueda académicos de libre acceso. La elección de este tipo de motores se debió al gran auge que están teniendo como alternativas a las bases bibliográficas tradicionales (*Scopus* y *Web of Science*). Se realizaron recuperaciones de registros para diferentes autores (algunos muy prolíficos y otros no tanto para observar las diferencias en los resultados), para este propósito se emplearon *Google Scholar* y *Microsoft Academic* como orígenes de datos. Para el caso de *Google Scholar* se construyó un *crawler* específico para la extracción de registros y para *Microsoft Academic* se empleó la *Academic Knowledge API*<sup>1</sup>. Ambos motores poseen diferencias sustanciales pero se intentó unificar un conjunto de datos básico con los atributos más importantes y en este sentido adaptar la visualización a dicho conjunto de datos.

Entre los atributos más importantes y que en la mayor parte de las bases de datos bibliográfica están presentes: el título de la publicación, año de publicación, autores, cantidad de citas recibidas (algunas bases de datos como los repositorios institucionales no incluyen este indicador), resumen, enlace al documento, enlace a los registros que citan la publicación y lugar de publicación. Como esta visualización se adaptó inicialmente a los motores de libre acceso mencionados, estos incluyen el tipo de archivo del registro (PDF entre los más comunes pero también existen elementos en formato .DOC y .DOCX, y en menor cantidad en formato .TXT y .PPT), en consecuencia se agregó también este elemento de información.

La idea de definir este conjunto de datos “básico” se realizó para poder independizar la visualización del origen de datos, en un intento de favorecer la integración de la herramienta con cualquier plataforma o base de datos que desee emplearla y que posea la tecnología adecuada (tecnología web y soporte para JavaScript como se verá más adelante).

Si bien no todos los campos de un registro pueden ser visualizados, la mayor parte son elementos textuales (título, resumen, enlaces) y pocos son datos numéricos, en un intento de cuantificar y convertir valores textuales en discretos se logró definir los atributos que formarían parte del gráfico quedando algunos como opciones de filtrado y el resto como elementos de in-

formación. En este sentido, los elementos del gráfico serán: año de publicación, número de citas recibidas, cantidad de autores por publicación y lugar de publicación. Como elementos de filtrado quedarían: el tipo de archivo (como variable adicional), el año de publicación, número de citas y lugar de publicación (estos tres últimos además de ser elementos propios del gráfico actuarán como filtros). El resto de atributos (título, resumen, autores, enlaces) se emplearán solo para mostrar información adicional de ese registro.

Teniendo en cuenta todo lo expuesto, se desarrolló un primer prototipo de visualización utilizando la librería D3.js (Data Driven Documents <https://d3js.org/>), Teller (2013); Zhu (2013); Nair et al. (2016).

Las posibilidades que entrega esta poderosa herramienta son inimaginables, si bien la curva de aprendizaje es empinada, la potencia y las capacidades de interacción que se pueden agregar en las distintas visualizaciones son enormes.

*ScatterPE* (Scatter Plot Extendido) es el nombre de la herramienta de visualización desarrollada en este trabajo, la misma recibe un archivo JSON como conjunto de datos de entrada, dicho archivo posee el conjunto de publicaciones de un investigador (este primer prototipo solo se ha utilizado para visualizar los resultados de un investigador individual, sin embargo no posee inconvenientes si lo que se desea evaluar son los registros de una institución, ya que lo único que aumentaría sería el número de registros a visualizar). Como la entrada es un archivo en formato JSON, esto permite que el origen de los datos pueda ser tanto de bases de datos tradicionales (*Scopus*, *WoS*) como de libre acceso (*Google Scholar*, *Microsoft Academic*, *Semantic Scholar*, etc.) de forma indistinta.

El formato del archivo JSON de entrada es el siguiente:

- **Title:** (string[500]) Título de la publicación.
- **Authors:** (string[500]) Lista de autores separados por “,” (coma).
- **Year:** (int) Año de publicación.
- **URL:** (string[400]) URL a la publicación ya sea al documento a texto completo o al repositorio donde está almacenado.
- **URLCites:** (string[400]) URL al conjunto de citas de dicha publicación.
- **Type:** (string[50]) Formato de la publicación, los formatos pueden ser: HTML, TXT, PDF, DOC, BOOK, PPT, XLS, PS u OTHERS.
- **CiteNumber:** (int) Número de citas.

<sup>1</sup>. <https://www.microsoft.com/en-us/research/project/academic-knowledge/>

- **JournalURL:** (string[250]) Para el caso de que los registros provengan de *Google Scholar*, dominio donde el documento ha sido localizado o en algunos casos el nombre de la Editorial, no se guarda el nombre de la revista pues gran parte de los registros no posee y en algunos casos aparece cortado. Para el caso de que la fuente de datos sea *Microsoft Academic* si se almacena el nombre de la revista, ya que la AK API si lo provee, aunque no todos los registros lo complimentan. La idea de este atributo es indicar un parámetro para agrupar un conjunto de registros, por consiguiente es indistinto si se utiliza el dominio, editorial o nombre de la revista, luego en la visualización se verá la utilidad del mismo.
- **Abstract:** (string[3000]) Resumen (abstract) de la publicación.
- **Source:** (string[5]) Abreviación del origen de los datos, por ejemplo: GS para Google Scholar o MA para Microsoft Academic.
- **Cites:** Es un array de objetos utilizado para almacenar las citas recibidas por la publicación en cuestión, si no es posible recuperar las citas (porque el motor académico no lo permite) o la publicación no posee citas, este atributo es vacío. Cada elemento posee los siguientes atributos: *Title*, *Authors*, *Year* y *URL*, los cuales poseen la misma especificación y significado que lo indicado para la publicación.

Es necesario aclarar que *ScatterPE* es una herramienta de visualización de información y no de recuperación de información, con lo cual la extracción de los registros de las distintas bases de datos queda fuera del alcance de este trabajo, es tarea del usuario proveer un conjunto de datos, por esta razón es que se independiza del origen de datos a partir del ingreso de un archivo JSON que puede tener cualquier origen siempre y cuando se respete el formato indicado anteriormente.

En *ScatterPE* los elementos se disponen del siguiente modo: en el eje vertical se encuentra la cantidad de coautores de las publicaciones, en el eje horizontal el año de publicación, cada publicación se representa por una burbuja donde el tamaño de las mismas indica la cantidad de citas recibidas, y por último el color de las burbujas indica el nombre de la revista/congreso donde fue publicado el registro bibliográfico. D3 posee un conjunto de opciones para crear e implementar animaciones e interacción con la visualización diseñada, algunas son sencillas de aplicar y otras requieren un esfuerzo considerable, además, éstas se pueden combinar para proporcionar un efecto mucho más

atractivo. Las animaciones están básicamente relacionadas con el cambio de posición de los elementos presentados o los efectos visuales para mostrarlos u ocultarlos. La interacción por su parte es muy similar a la animación, solo que los cambios en el contexto visual son operaciones que están bajo el control del usuario, es decir, es el usuario el que inicia la interacción con la herramienta mediante una acción, como hacer *clic* en un botón, como arrastrar el *mouse*, como hacer *scroll*, entre tantas otras.

El prototipo diseñado ofrece un conjunto de seis opciones de interacción entre las que se incluye: la posibilidad de filtrar por el origen de la publicación, este elemento se representa por un recuadro con un color distinto para cada ítem. Al hacer *clic* sobre el recuadro de color de un elemento, la herramienta oculta las burbujas (publicaciones) que poseen dicho valor o categoría, una vez hecho esto la categoría se torna de color gris indicando que los elementos correspondientes están ocultos. Al hacer *clic* nuevamente sobre el recuadro de color gris, la herramienta muestra las burbujas que poseen dicho valor o categoría, una vez hecho esto la categoría retoma su color original indicando que los elementos correspondientes están visibles.

La segunda interacción que ofrece tiene que ver con el filtrado del formato de la publicación, es decir, con el tipo de archivo del recurso visualizado, la opción es un *OptionButton* que puede tomar cualquier de los siguientes valores: ALL, HTML, TXT, PDF, DOC, BOOK, PPT, XLS, PS u OTHERS, en la categoría OTHERS entran aquellos registros que no tienen indicado el formato o tipo de archivo, o aquellos que no han podido ser identificados como uno de los formatos mencionados.

Como se observa en la figura 4, la opción por defecto que muestra todos los elementos sin filtrar es "ALL", al hacer clic en cualquiera de los otros valores, las burbujas desaparecen o aparecen según se elija un tipo de archivo u otro.

Otra interacción es la posibilidad de filtrar los registros por año de publicación, a tal efecto el filtro es un control de tipo *Slider*. Posee dos controles *slider*, uno para indicar el inicio del intervalo *From year* y otro para indicar el final del intervalo *To year*. Ambos filtros funcionan de forma independiente y las burbujas nuevamente, aparecen o desaparecen según posean un valor de año de publicación que esté dentro del rango de los filtros de año elegido.

La cuarta interacción es la posibilidad de visualizar el detalle de la publicación al pasar el mouse por encima de una burbuja, el detalle mostrado incluye: el título de la publicación (Title), listado de autores (Authors), año de publicación (Year), número de citas (Cite Number), tipo de archivo (Type), lugar de publicación/nombre de la revista/categoría (Publication source) y Resumen (Abstract). Al quitar el mouse del radio de la burbuja el detalle mostrado desaparece (ver figura 5).

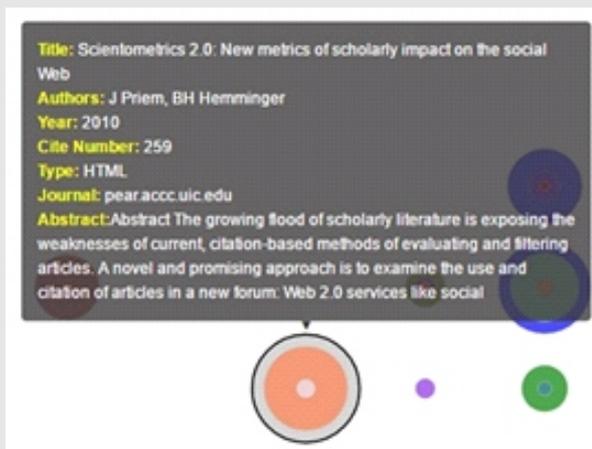


Fig. 5 Detalle de una publicación.

La quinta interacción permite filtrar las burbujas de acuerdo a la cantidad de citas recibidas, así pues existen dos controles de tipo *Slider*. Un control para indicar el inicio del intervalo *From #Cites* y otro para indicar el final del intervalo *To #Cites*. Ambos filtros funcionan de forma independiente y las burbujas aparecen o desaparecen según posean un valor de cantidad de citas recibidas que esté dentro del rango los filtros de cantidad de citas elegido.

La última interacción es la posibilidad de acceder al recurso publicado por medio del link que se almacenó al momento de hacer la recuperación o importación de datos, para ello es necesario hacer doble click en cualquier parte de la burbuja, esto abrirá una nueva pestaña con el recurso solicitado.

La superposición de las burbujas se da cuando más de una publicación comparten el mismo año de publicación y la misma cantidad de coautores, sin embargo esta superposición es tal que las publicaciones con mayor número de citas se ubican al fondo mientras

que las de menor cantidad de citas se ubican por encima.

Con las interacciones y la representación propuestas se observa que es posible obtener diferentes vistas desde una única visualización, las múltiples opciones de filtrado de los elementos logran este objetivo.

## Trabajos futuros

El desarrollo que se presenta en este trabajo no es una herramienta finalizada, al contrario, es un prototipo inicial con mucho potencial y cuestiones a mejorar, la idea es ir avanzando y dotándola de mayores prestaciones. Un agregado inicial será la incorporación de indicadores bibliométricos, puesto que se cuenta con la cantidad de citas por publicación, estos serán fácilmente calculados. Otro agregado que aportaría más información sería la incorporación de los datos de afiliación de los autores (en caso que sea posible recuperarlos desde la fuente de datos), en este sentido se podrían identificar redes de colaboración y otras relaciones entre instituciones.

Se plantea la posibilidad de incorporar la capacidad de filtrar trabajos por palabras claves o frases, de este modo solo se visualizarían los trabajos que cumplan la existencia de estos términos tanto en el título de la publicación como en el resumen.

La herramienta fue probada por un grupo de investigadores en un entorno preparado para tal fin, con varios conjuntos de datos para observar las diferencias entre investigadores muy productivos y otros menos productivos. Luego de las pruebas y de una revisión inicial, los investigadores que la probaron destacaron que una mejora sustancial sería brindar la opción de comparar dos conjuntos de datos distintos, ya sea del mismo investigador provenientes de fuentes distintas o comparando dos investigadores distintos.

Y por último, ya que se cuenta con información del título y resumen de las publicaciones, sería interesante mostrar las relaciones que se pueden hallar entre los registros de un conjunto de datos, esto serviría para revelar registros semánticamente relacionados o que comparten algún término o frase.

## Conclusiones

Se ha presentado un primer prototipo de una herramienta de visualización de información de publicaciones científico-académicas que puede resultar bastante útil al momento de realizar un análisis bibliométrico. *ScatterPE* viene a cubrir una falta notable en esta temática puesto que no solo se independiza del origen de datos, sino que ofrece un conjunto de interacciones al usuario que permitirán en mayor o menor medida tener una visión distinta y ampliada de los datos que se puedan recolectar a partir del análisis llevado a cabo.

Es necesario destacar también que al ser una herramienta web y desarrollada con librerías de acceso li-

bre, puede ser fácilmente incorporada en cualquier plataforma web o sistema de evaluación o base de datos bibliográfica. La integración no demanda recursos adicionales ni un profundo conocimiento más que el de conocer programación web.

Las pruebas iniciales sobre la herramienta desarrollada arrojaron resultados muy alentadores, destacando por un lado que el conjunto de interacciones provisto aporta distintas visiones en una misma representación y por otro lado la velocidad con la que procesa el conjunto de datos.

Por último y no menos importante, resulta necesario destacar el gran trabajo que queda por delante para aumentar y mejorar las prestaciones de la herramienta aquí presentada.

## Referencias Bibliográficas

- Bornmann, L., Hug, S. & Marewski, J. N.** (2018) Bibliometrics-based heuristics: What is their definition and how can they be studied?. *arXiv preprint arXiv:1810.13005*.
- Bostock, M.** (2019) *Observable*. [En línea] Consultado en Noviembre/2019: <https://observablehq.com/@d3/scatterplot>
- Card, S., Mackinlay, J. & Shneiderman, B.** (1999) *Readings in Information Visualization: Using Vision to Think*. San Francisco: Morgan-Kaufmann.
- Chen, Q., Yue, X., Plantaz, X., Chen, Y., Shi, C., Pong, T. y Qu, H.** (2018) Viseq: Visual analytics of learning sequence in massive open online courses. *IEEE transactions on visualization and computer graphics*.
- Harzing, A.-W.** (2017) Google Scholar is a serious alternative to Web of Science. *LSE Impact of Social Sciences Blog*.
- Harzing, A.-W. & Alakangas, S.** (2017) Microsoft Academic is one year old: The Phoenix is ready to leave the nest. *Scientometrics*, 112(3), pp. 1887--1894.
- Keim, D. A.** (2002) Information visualization and visual data mining. *IEEE Transactions on Visualization and Computer Graphics*, 7(1).
- Köpp, W. & Weinkauff, T.** (2018) Temporal treemaps: Static visualization of evolving trees. *IEEE transactions on visualization and computer graphics*, 25(1), pp. 534--543.
- Kosara, R., Hauser, H. & Gresh, D.** (2003) *An interaction view on information visualization*. s.l., EUROGRAPHICS, pp. 123-137.
- Lacey, M.** (2017) *Statistical topics*. [En línea] Consultado en Noviembre/2019: <http://www.stat.yale.edu/Courses/1997-98/101/scatter.htm>
- Liao, H., Tang, M., Luo, L., Li, C., Chiclana, F. y Zeng, X.** (2018) A bibliometric analysis and visualization of medical big data research. *Sustainability*, 10(1), p. 166.

**Nair, L., Shetty, S. & Shetty, S.** (2016) Interactive visual analytics on Big Data: Tableau vs D3.js. *Journal of e-Learning and Knowledge Society*, 12(4).

**Teller, S.** (2013) *Data Visualization with D3.js*. s.l.:Packt Publishing.

**Yeung, A. W. K.** (2018) Data visualization by alluvial diagrams for bibliometric reports, systematic reviews and meta-analyses. *Current Science*, 115(10), pp. 1942-1947.

**Zhu, N. Q.** (2013) *Data Visualization with D3.js Cookbook*. s.l.:Packt Publishing.

Este artículo se realizó en el Laboratorio de Virtualización de la Facultad de Ingeniería de la Universidad Nacional de Jujuy (UNJu), durante el primer semestre del año 2019 en el marco del proyecto de investigación bianual D/B029 de la Secretaría de Ciencia y Técnica y Estudios Regionales de la UNJu (SeCTER-UNJU) denominado "Aplicación de técnicas de Inteligencia Artificial para evaluar la producción científico-académica de investigadores de Universidades públicas del Noroeste Argentino".

### José Federico Medrano

Doctor en Informática y Automática por la Universidad de Salamanca, España (2017), Máster en Sistemas Inteligentes por la Universidad de Salamanca, España (2011) e Ingeniero Informático por la Universidad Nacional de Jujuy (UNJu), Argentina (2008). Docente de las carreras Ingeniería Informática, Licenciatura en Sistemas y Analista Programador Universitario de la Facultad de Ingeniería de la UNJu (FI UNJu). Coordinador del gabinete de capacitación de RRHH de la carrera Ingeniería Informática de la Universidad Católica de Santiago del Estero sede San Salvador de Jujuy. Director del grupo de investigación VRAIn de la FI UNJu. Director de proyectos de Investigación acreditados por la SeCTER UNJu y de vinculación de la SPU. Se especializa en el análisis y diseño de sistemas, empleo de bases de datos bibliográficas de libre acceso, cibermetría, visualización y recuperación de información, y aprendizaje automático.

E-mail: [jfmedrano@fi.unju.edu.ar](mailto:jfmedrano@fi.unju.edu.ar)



# cet

REVISTA DE CIENCIAS EXACTAS E INGENIERIA  
FACULTAD DE CIENCIAS EXACTAS Y TECNOLOGÍA