

## Descubriendo críticas al acceso abierto mediante la visualización de textos con *Voyant Tools*

### Uncovering open access criticisms by visualizing texts with *Voyant Tools*

Nancy Sánchez Tarragó<sup>1\*</sup> <https://orcid.org/0000-0002-5114-6072>

<sup>1</sup>Universidade Federal de Rio Grande do Norte. Departamento de Ciência da Informação. Brasil.

\*Correspondencia: [nancita1973@gmail.com](mailto:nancita1973@gmail.com)

#### RESUMEN

*Voyant Tools* es un entorno web de código abierto diseñado para facilitar el ciclo hermenéutico por medio de la navegación entre diferentes niveles de lectura, análisis y visualización interactiva de textos (artículos científicos, páginas web, mensajes de Twitter, obras literarias, etc.). El objetivo de este artículo es ejemplificar cómo podemos usar esta herramienta para la exploración y análisis de textos científicos, que permitan descubrir temas y asociaciones entre documentos. Se utiliza como *corpus* para el análisis 33 documentos científicos que cuestionan y problematizan los movimientos de acceso abierto y ciencia abierta. Por medio de diferentes herramientas analíticas y de visualización emergen términos (*open source*, *APC*, *impact factor*, *prestige*, *exclusión*, *colonialism*, *coloniality*) que, como rastros de “migajas de pan”, permiten identificar autores, documentos y problematizaciones relacionadas con el acceso abierto y la ciencia abierta. Se concluye que *Voyant Tools* es una herramienta versátil e intuitiva con diversas potencialidades para el análisis y el descubrimiento de información en los textos. Con respecto al tema explorado, se identifican críticas y problematizaciones relacionadas con el énfasis en el mercado y la aceleración

capitalista por detrás de políticas e iniciativas de acceso abierto y ciencia abierta, el auge de revistas de acceso abierto comerciales que cobran por procesamiento de artículos (APC) y la aparición de nuevos colonialismos y epistemicidios cuando se combinan acceso abierto, mercantilización del conocimiento y uso distorsionado del factor de impacto.

**Palabras clave:** *Voyant Tools*; visualización de textos; acceso abierto; ciencia abierta; críticas.

## ABSTRACT

*Voyant Tools* is an open-source web environment designed to facilitate the hermeneutical cycle by navigating between different levels of reading, analysis, and interactive visualization of texts (scientific articles, web pages, Twitter messages, literary works). The objective of this article is to exemplify how we can use this tool for the exploration and analysis of scientific texts, discovering themes and associations between documents. 33 scientific documents that question and problematize the open access and open science movements are used as a corpus for analysis. Through different analytical and visualization tools, emerge terms (open source, APC, impact factor, prestige, exclusion, colonialism, colonality) that as traces of "breadcrumbs" serve to make new explorations and visualizations that allow to identify authors, documents and problems related to open access and open science. It is concluded that *Voyant Tools* is a versatile and intuitive tool with various potentialities for the analysis and discovery of information in texts. Regarding the topic explored, criticisms and problematizations related to the emphasis on the market and capitalist acceleration behind open access and open science policies and initiatives are identified, with the rise of the commercial open access journals' business model based on article processing charge (APC) and the appearance of new colonialism and epistemicide when open access, commodification of knowledge and distorted use of the impact factor are combined.

**Key words:** *Voyant Tools*; text visualization; open access; open science; critique.

Recibido: 01/01/2021

Aceptado: 16/03/2021

## Introducción

Para el ser humano, los procesos de representación son esenciales para la creación, el almacenamiento y la transmisión del conocimiento. Representar es el acto de usar elementos simbólicos (palabras, figuras, imágenes, dibujos, mímicas, esquemas, entre otros) para substituir un objeto, idea o hecho.<sup>(1)</sup> Pinturas rupestres, obras literarias, artículos científicos, diseños técnicos, grafitis, metadados, infográficos, nubes de palabras; son todos ejemplos de representaciones que el ser humano ha creado para representar tanto la realidad como los mundos imaginarios.

*Alvarenga*<sup>(2)</sup> apunta que las representaciones pueden ser primarias, aquellas que realizan los autores en el momento de expresar sus pensamientos, y secundarias, las que realizan los profesionales de la información a partir de aquellas representaciones primarias. En este caso se encuentran los metadatos para describir documentos (encabezamientos de materias, términos de indexación, registros bibliográficos, resúmenes), pero también, por ejemplo, las representaciones gráficas y tabulares que sintetizan visualmente los resultados de estudios bibliométricos.

Una característica de las representaciones es que siempre implican un proceso de reducción semántica; un proceso de síntesis y abstracción que, sin embargo, es capaz de revelar y ampliar nuevos significados. Esta posibilidad de traer nuevos significados e interpretaciones está en la base de la visualización de la información, un área interdisciplinar relativamente nueva que surgió en la década de 1990, en la cual las representaciones gráficas (palabras, imágenes, números) se potencializan con tecnologías computacionales que permiten el procesamiento y análisis estadístico de grandes volúmenes de datos. La visualización, mediante ese poder reductor, ayuda a percibir a simple vista

ciertos datos y fenómenos de una realidad compleja y multidimensional que de otro modo pasarían desapercibidos. <sup>(3)</sup>

Por otra parte, los últimos años han visto crecer el número de aplicaciones de la visualización para el análisis de textos. En este caso, se pretende “representar sus características importantes de forma más compacta y eficiente, a menudo mediante la visualización. Las visualizaciones son transformaciones de textos que tienden a reducir la cantidad de información presentada, pero al servicio de llamar la atención sobre algún aspecto significativo”.<sup>(4)</sup>

Diversos son los tipos de textos que pueden someterse a estos análisis: obras literarias, cartas, entrevistas, textos científicos, blogs, publicaciones de redes sociales. También diversas aplicaciones computacionales han sido desarrolladas para este objetivo. Una de ellas es *Voyant Tools*, un entorno web de código abierto, para lectura, análisis y visualización de textos, surgido como “proyecto académico para facilitar la lectura y las prácticas interpretativas para estudiantes y académicos de humanidades digitales, así como para el público en general”.<sup>(5)</sup> Constituye un conjunto integrado y compacto de herramientas analíticas y de visualización.

Los resultados de búsquedas bibliográficas en diversas bases de datos sugieren que su uso es incipiente entre los profesionales de las ciencias de la información, a pesar de sus potencialidades. Por eso, este artículo tiene un enfoque metodológico y exploratorio, pues su objetivo es ejemplificar los procedimientos interactivos de uso de *Voyant Tools* para el análisis y la visualización de textos científicos, que permitan descubrir temas y asociaciones entre documentos. Profesionales de la información de diversos ámbitos podrían valerse de estas herramientas de análisis de textos para, en unos casos, apoyar sus investigaciones; en otros, para la elaboración de servicios de información especializados.

Los textos aquí usados para mostrar las posibilidades de análisis de *Voyant Tools* son textos que cuestionan y problematizan las iniciativas, políticas y modelos de comunicación científica relacionados con el acceso abierto. Este movimiento aboga por el acceso irrestricto e inmediato a la literatura científica. Varios trabajos han discutido sus beneficios académicos, sociales y económicos.<sup>(6)</sup> No obstante, muchas veces cuando se habla del acceso abierto el tono casi siempre es optimista y triunfante, asumiendo la apertura a los datos y la información como indispensable, neutral, universal, un imperativo moral que hay que alcanzar a toda costa y que sin dudas conducirá al progreso y al desarrollo, entendido este último casi siempre en su vertiente económica.

Sin embargo, un creciente número de trabajos alertan para nuevas formas de privatización del conocimiento, colonialismo, eurocentrismo y exclusión,<sup>(7)</sup> lo que evidencia las complejidades de un fenómeno multidimensional con aristas relacionadas con la economía de las publicaciones, la política científica, los sistemas de evaluación y recompensa académica, los regímenes de propiedad intelectual, las tecnologías de información, en un contexto de hegemonía del capitalismo neoliberal. El estudio de esta temática se incluye dentro del proyecto de investigación *Conhecimento e poder nos discursos em organização e comunicação do conhecimento: um mapeamento desde a Ciência da Informação*, del cual soy coordinadora.

Por tanto, pretendo mostrar que *Voyant Tools* es útil para explorar interactivamente el contenido de estos documentos e identificar temas que lleven a descubrir autores, documentos, críticas y problematizaciones sobre el acceso abierto, en un ciclo interpretativo de lectura, análisis y visualización.

## Referencial teórico

### Visualización de la información

Una visualización es una representación por medio de imágenes, gráficos o cartografías, cuyo propósito esencial es la simplificación o reducción de los contenidos para permitir la comprensión de la idea general.<sup>(8)</sup> Aunque estamos familiarizados con mapas, gráficos simples y cuadros, la aplicación creciente

de la computación hace que hoy la visualización de la información sea un área interdisciplinar que permite la representación de diversos objetos, incluyendo diferentes tipos de datos, algoritmos, resultados de cálculos, procesos y muchos otros componentes a solicitud del usuario.<sup>(9)</sup> Por tanto, la visualización es más que simples gráficos por computadora, pues su campo abarca aspectos de muchas otras disciplinas, incluida Interacción Humano-Computador, Psicología Perceptiva, Bases de Datos, Estadísticas, Minería de Datos, Ciencias Cognitivas, Ciencias de la Computación, Cartografía, Diseño Gráfico, Artes Gráficas, Ergonomía Cognitiva, entre otras.<sup>(8,10,11)</sup>

Presupone que los seres humanos no son receptores pasivos de estímulos, sino que amplían su conocimiento en la medida en que interactúan y realizan asociaciones con representaciones textuales o visuales. Estas últimas son muy importantes para el proceso de cognición, ya que expresan una conexión innata entre la visión y los procesos de raciocinio humano y de toma de decisiones.<sup>(11)</sup> Como apuntan *Ward, Gristein e Keim* “[...]Una única imagen puede contener una riqueza de informaciones y puede ser procesada mucho más rápido que una página con palabras escritas. Eso ocurre porque la interpretación de las imágenes se realiza en paralelo dentro del sistema de percepción humana, mientras que la velocidad de análisis de un texto es limitada por el proceso de lectura secuencial”.<sup>(9)</sup>

La visualización de la información puede ser aplicada a diversos ámbitos (sociales, económicos, políticos y científicos) y en diferentes disciplinas científicas. En las Ciencias de la Información se ha utilizado para una variedad de propósitos, entre ellos, para explorar la estructura de campos científicos, como complemento de los estudios métricos; en recuperación de información; en interfaces de bibliotecas digitales y catálogos en línea; en la representación y análisis de espacios documentales y dominios de conocimiento.<sup>(10)</sup>

Hoy día, las técnicas de visualización de información combinan minería o análisis automatizado de datos (estructurados, semiestructurados y no estructurados) con visualizaciones interactivas, donde el usuario o analista

interactúa con los resultados en función de sus objetivos. Este análisis visual comprende un ciclo iterativo: colecta de datos, procesamiento, creación de visualizaciones, interacción y toma de decisiones.<sup>(8)</sup> Una parte de ese proceso ocurre en la computadora (procesamiento, transformación, cálculos algorítmicos), pero otra parte es responsabilidad del usuario, el analista humano que va a observar la visualización, interpretarla y si juzga necesario, va a alterar los parámetros para realizar nuevas actualizaciones en las visualizaciones, lo que reiniciaría el ciclo.

Una buena visualización de la información permite representar grandes volúmenes de información a pequeña escala; menor tiempo para la búsqueda de información; alcanzar un buen conocimiento de estructuras de datos complejas; identificar relaciones que de otro modo pasarían desapercibidas; observar un conjunto de datos desde diferentes perspectivas; estimular la formulación de hipótesis, y provocar debate y discusión sobre el tema o fenómeno representado.<sup>(12)</sup>

Es necesario destacar que, en cualquier tipo de visualización, las habilidades y las limitaciones de la percepción humana son muy importantes y estas varían de individuo en individuo, influenciadas incluso por su familiarización con el dominio de donde provienen los datos. No es posible pensar en la visualización como algo “dado”, que existe independientemente de nuestras capacidades de percepción/interpretación y construcción de significados.

### **Minería/análisis y visualización de textos**

Actualmente, es común la existencia de grandes colecciones digitales de textos, sean artículos científicos, libros, periódicos, mensajes de correo electrónicos, post de redes sociales y archivos de distinta naturaleza. La tarea de descubrir información/conocimiento en esos volúmenes de datos es desafiadora. Al igual que con otros tipos de datos, técnicas de minería y visualización de textos han resultado de gran provecho.

Para las ciencias de la computación, la minería o análisis computacional de textos se considera generalmente una subárea de la minería de datos; su propósito es descubrir conocimiento (patrones) en datos no estructurados o semiestructurados en lenguaje natural.<sup>(13,14)</sup> Sin embargo, para las ciencias sociales y las humanidades, la minería de textos bebe de la tradición interdisciplinar de la lingüística de *corpus* y la lingüística computacional, así como del análisis de redes sociales, para respaldar interrogantes de investigación en estas áreas.<sup>(13)</sup> Los enfoques cuantitativos para estudio de textos tienen una historia que se remonta al siglo XIX, sin embargo, no es hasta la primera década del siglo XXI, con la digitalización masiva de los textos y la implementación de nuevas y cada vez más sofisticadas herramientas, que estas aplicaciones son cada vez más ubicuas;<sup>(4)</sup> ellas complementan y potencializan el tradicional análisis científico que se hace a partir de lectura, anotaciones, síntesis y asociaciones. Por tanto, la minería y visualización de textos es un área interdisciplinar y sus aplicaciones van desde las ciencias biomédicas hasta el campo vasto que hoy se denomina humanidades digitales.

Al igual que en la minería de datos estructurados, la minería de textos requiere un enfoque semiautomático pues antes de la etapa de descubrimiento de conocimiento (que será realizada por el analista humano) es necesario procesar de forma automatizada grandes colecciones documentales (*corpus* documental) y transformarlas en un formato que facilite su comprensión y análisis.<sup>(15)</sup> Para esto se utilizan diversas técnicas de recuperación de información, procesamiento del lenguaje natural y análisis estadístico multivariado, que permiten, por ejemplo, análisis lexical, análisis de *clusters*, análisis de correspondencia, categorizaciones, así como diversos tipos de visualizaciones interactivas.

Una fase importante de este proceso es la interpretación de esas visualizaciones. *Sinclair* y *Rockwell*<sup>(4)</sup> se refieren a un ciclo hermenéutico o interpretativo que ocurre entre la lectura, el análisis y la visualización: “[...] leemos textos que disfrutamos, luego los exploramos y estudiamos con



herramientas analíticas e interfaces de visualización, lo que nos lleva de nuevo a releer los textos de manera diferente. Esto es lo que llamamos un ciclo interpretativo ágil.”

Los autores indican que hay dos tipos de interrogantes que justifican el uso de estas técnicas de análisis y visualización:<sup>(4)</sup>

1. ¿Cómo pueden ayudarnos a identificar y estudiar aspectos interesantes que no habíamos notado antes en textos con los que ya estamos familiarizados, o aspectos ya percibidos para los cuales no teníamos medios razonables de análisis?
2. ¿Cómo pueden ayudarnos a identificar y comprender textos con los que no estamos familiarizados o que no tenemos condiciones razonables de leer?

En relación con estas interrogantes, *Sinclair* y *Rockwell* argumentan que a través de la amplia gama de representaciones ofrecidas por estas herramientas podremos explorar características lingüísticas y semánticas, y así, producir nuevas representaciones y asociaciones; esto nos puede ayudar a fortalecer las intuiciones que ya teníamos o a generar perspectivas completamente nuevas. Por otra parte, estos sistemas pueden ayudar a extender la capacidad de lectura y comprensión humana, usualmente limitadas, especialmente si se trata de grandes colecciones de textos, y a la misma vez, ayudar a identificar lo que podríamos realmente querer leer en estas colecciones.<sup>(4)</sup>

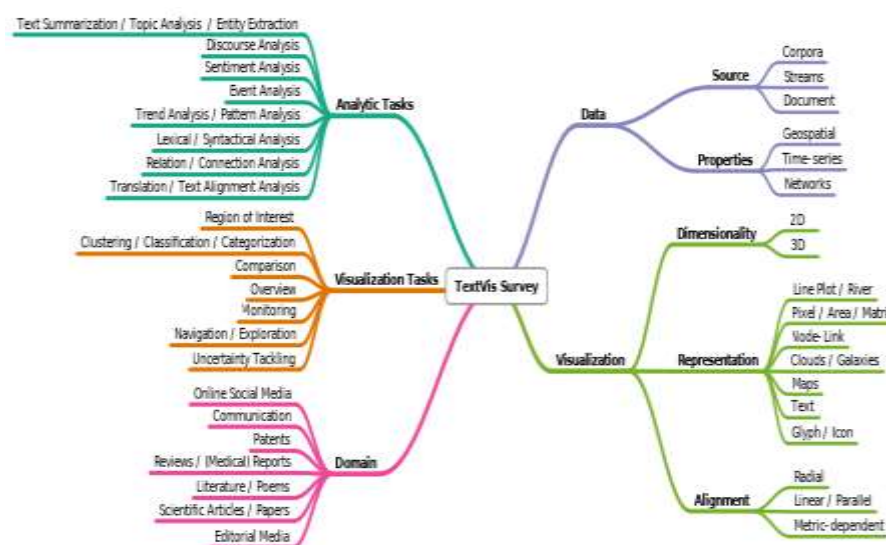
En este contexto, las técnicas de visualización se aplican a tres grandes tipos de textos:<sup>(16)</sup>

*Documentos:* por ejemplo, un artículo de una revista o de un periódico, una página web. El análisis visual generalmente tiene como objetivo ilustrar cómo se relacionan varios documentos, o resumir el contenido o sus características lingüísticas para facilitar una comprensión y comparación eficaz de varios documentos.

*Corpus*: se refiere a una colección de documentos. Visualizaciones diseñadas para representar el *corpus* generalmente se enfocan en revelar estadísticas que permitan identificar, por ejemplo, los temas principales del conjunto de datos.

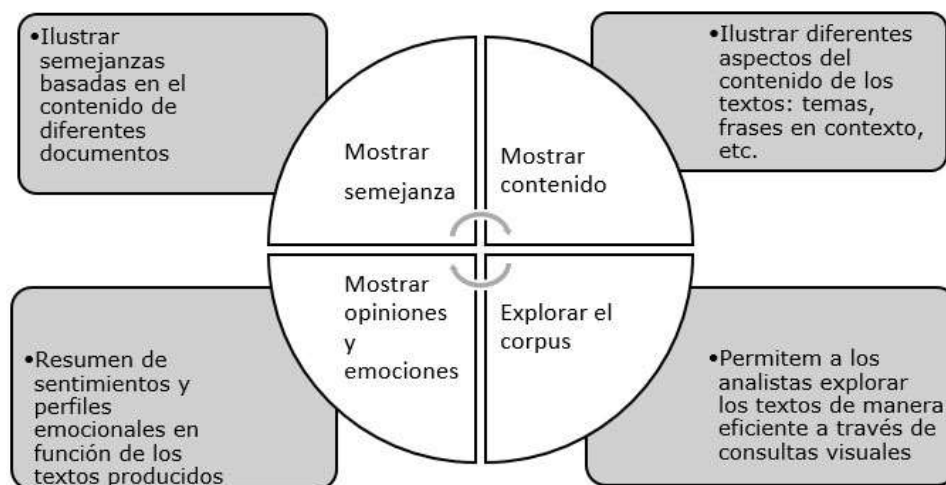
*Cadenas de caracteres (strings)*: cualquier tipo de texto, como los mensajes en Twitter. La visualización ayuda, por ejemplo, a mostrar la tendencia general de los datos a lo largo del tiempo.

A partir de estos tipos de datos textuales se han desarrollado varias técnicas de análisis y visualización. El grupo ISOVIS, de la Linnaeus University, Suecia, mantiene una página interactiva con una compilación de más de 400 técnicas de visualización de textos (<https://textvis.lnu.se/>). Kucher y Kerren,<sup>(17)</sup> pertenecientes al grupo ISOVIS, han realizado una taxonomía de las técnicas (Fig. 1), donde clasifican las principales tareas analíticas y de visualización, además de los dominios de aplicación más frecuentes, las fuentes y las propiedades de los datos, así como las características de las visualizaciones. En sentido general, las técnicas pueden ser clasificadas en cuatro grandes categorías, según sus propósitos (Fig. 2): mostrar el contenido de los documentos; explorar el corpus; mostrar la semejanza y diferencia entre los vocabularios de los documentos; y mostrar opiniones y sentimientos.



Fuente: Kucher y Kerren.<sup>(17)</sup>

Fig. 1 - Taxonomía de técnicas de análisis y visualización de textos.



Fuente: elaborada a partir de Cao y Cui.<sup>(16)</sup>

**Fig. 2** - Categorización de técnicas de análisis y visualización de textos según sus propósitos.

Las herramientas para realizar estos análisis también son diversas; algunas son *softwares* propietarios, mientras que otras son de código abierto y pueden ser libremente accesibles en formato web o de escritorio.<sup>(18,19)</sup> Un ejemplo de herramientas de código abierto es *Voyant Tools* (<http://www.voyant-tools.org>), considerado como un proyecto académico en continuo desarrollo, liderado por los profesores e investigadores de humanidades digitales *Stéfan Sinclair*, de la McGill University, y *Geoffrey Rockwell*, de la University of Alberta, ambas instituciones con sede en Canadá. *Voyant Tools* es un ambiente web para leer, analizar y visualizar textos. Posee además una versión de escritorio. Los autores también mantienen el sitio *Hermeneuti.ca* (<http://hermeneuti.ca/>), que incluye versiones en línea de los capítulos prácticos del libro de su autoría *Hermeneutica: computer-assisted interpretation in the humanities*<sup>(20)</sup> con paneles dinámicos de estudios de casos con *Voyant Tools*. La herramienta también posee un sitio web con una detallada documentación (<https://voyant-tools.org/docs/#!/guide>). Otro tutorial puede ser consultado en *Gutiérrez de la Torre*.<sup>(21)</sup>

*Voyant Tools* ofrece un sinnúmero de herramientas analíticas y de visualización que permiten también alternar entre ellas. Sus creadores aclaran que *Voyant Tools* no va a dar respuestas definitivas para las preguntas de investigación, sino que está diseñado para ayudar en las prácticas exploratorias e interpretativas de los textos, lo que facilita el ciclo hermenéutico por medio de la navegación entre diferentes niveles de lectura, análisis y visualización de los textos. O sea, permite tanto la llamada lectura cercana (la lectura del texto y sus frases en contexto) como la lectura distante (una visión panorámica y de zoom facilitada por la minería de textos y la visualización interactiva).

Una búsqueda exploratoria en diferentes bases de datos (*Directory of Open Access Journals*, *Emerald Insight*, *Library & Information Science Abstracts* (LISA), *Library, Information Science & Technology Abstracts* (LISTA), *Redalyc*, *Scopus* y *Web of Science*), realizada en diciembre del año 2020 por el término *Voyant Tools*, arrojó relativamente pocos resultados, donde predominaron los trabajos en inglés, de autores de instituciones de los Estados Unidos y publicados entre los años 2019 y 2020. Esto sugiere, por una parte, un aumento reciente de interés en la herramienta, y por otra, tal vez una escasez de trabajos en otros idiomas y contextos. En el sitio de *Voyant Tools* se mantiene una galería de ejemplos con investigaciones que usan la herramienta (<http://docs.voyant-tools.org/about/examples-gallery/>), aplicados a diversos tipos de documentos.

Una selección de los trabajos recientemente publicados que utilizan *Voyant Tools* muestra diferentes propósitos y contextos de aplicación: análisis de la autodenominación y autoclasificación de la homosexualidad masculina en Brasil;<sup>(22)</sup> descubrimiento de los patrones de términos y temas en la producción científica de una facultad universitaria;<sup>(23)</sup> análisis y comparación de los textos de misiones de universidades;<sup>(24)</sup> e identificación del lenguaje jurídico archivístico colombiano en diferentes períodos.<sup>(25)</sup> El trabajo de *Miller*<sup>(26)</sup> es también un ejemplo interesante de evaluación de las capacidades de *Voyant Tools* para trabajar con colecciones en humanidades digitales.

## Enfoques críticos y problematizaciones sobre el acceso abierto y la ciencia abierta

Hace casi 20 años, la declaración de la Iniciativa de Budapest formalizaba el Movimiento de Acceso Abierto a la Información, donde establecía como su foco los artículos científicos y definía como acceso abierto, “su disponibilidad gratuita en la Internet pública, lo que permite a cualquier usuario leer, descargar, copiar, distribuir, imprimir, buscar o enlazar a los textos completos de estos artículos, rastrearlos para indexarlos, pasarlos como datos al software o utilizarlos para cualquier otro fin lícito, sin barreras financieras, legales o técnicas que no sean las inseparables del acceso a Internet”.<sup>(27)</sup> Fueron establecidas como sus principales rutas el depósito o autoarchivo en repositorios y la publicación en revistas de acceso abierto.

Diversos factores y prácticas se reconocen en su genealogía: prácticas preexistentes en algunas comunidades científicas de compartimento público de *preprints* como, por ejemplo, la comunidad de física de altas energías; los movimientos de *software* libre y código abierto; el recrudescimiento de las restricciones de derecho de autor -que dejaron a los autores sin derechos sobre sus propias obras-; el auge de sistemas de evaluación y recompensa académica basados en indicadores bibliométricos como el factor de impacto; la demanda inflexible de revistas científicas por su asociación con prestigio y promoción de carreras; la aparición de oligopolios editoriales con enorme poder de fijación de precios y el aumento abusivo del precio de suscripción de las revistas científicas, que llegó a ser denominado como “crisis de las revistas”.<sup>(28,29)</sup>

En ese contexto, unos partidarios del acceso abierto enfatizan la necesidad de proporcionar acceso gratuito a los trabajos científicos (por ejemplo, a través de repositorios), mientras que otros (a la manera de la tradición del *software* de código abierto), se centran, no en la gratuidad del producto, sino en la libertad de uso, por medio de licencias *Creative Commons*.<sup>(29)</sup> A pesar de la disparidad de motivaciones, nociones y valores asociados al acceso abierto, sin dudas, ganó la atención de bibliotecarios, editores, científicos y gestores de

todo el mundo. Se implantaron numerosos repositorios temáticos e institucionales; proliferaron las declaraciones de organismos regionales e internacionales; financistas, instituciones y gobiernos implantaron políticas para presionar las prácticas de depósito y la publicación en revistas de acceso abierto; se crearon y diversificaron las revistas de acceso abierto en todas las áreas científicas.

En Latinoamérica, donde ya existía una tradición de revistas electrónicas gratuitas, subvencionadas por universidades públicas y gobiernos se comenzó a reconocer que estábamos a la delantera del Movimiento de Acceso Abierto.<sup>(30,31)</sup> Mientras, en otras partes del mundo, se crearon nuevas revistas de acceso abierto y las revistas comerciales comenzaron a experimentar con modelos de negocios que, para ofrecer acceso gratuito a los lectores, transferían el costo a los autores a través de los llamados cargos por procesamiento de artículos (APC, siglas en inglés). Actualmente, aunque en número de títulos predominan aún las revistas de acceso abierto que no cobran ni a autores ni a lectores (llamadas revistas Diamante), fundamentalmente publicadas por universidades y órganos públicos, las revistas que cobran APC (revistas Oro) publican alrededor del 60 % del total de artículos en acceso abierto,<sup>(32)</sup> lo que evidencia la utilización creciente de este modelo por parte de las principales editoriales comerciales.

Los cuestionamientos, críticas y problematizaciones sobre el acceso abierto surgieron desde su formalización, pero sobre todo, desde que las editoras de revistas en acceso abierto BioMed Central y *Public Library Online* estrenaron a principios de los años 2000 su modelo basado en APC,<sup>(33)</sup> lo que implicaba posibles conflictos de interés financiero y exclusión de autores que no tendrían condiciones para costear sus publicaciones. Desde entonces, las problematizaciones se han hecho más complejas y se han extendido a la ciencia abierta, como término sombrilla que reúne prácticas heterogéneas relacionadas con la apertura de todo el ciclo de investigación y comunicación. ¿Es el acceso abierto una exigencia moral, indispensable para todo tipo de información? ¿Es suficiente que el conocimiento esté ‘abierto’ para que sea utilizado? ¿Será el acceso abierto un (otro) instrumento para mantener a los países del Sur,

aquellos con una historia de colonialismo y desigualdades, sometidos a relaciones de subordinación epistémica? ¿El acceso abierto y la ciencia abierta son instrumentos de democratización del conocimiento, o instrumentos del capitalismo neoliberal que ahonda las brechas sociales y económicas? ¿Implicará el acceso abierto una real transformación de la comunicación científica mientras el factor de impacto y las revistas, como moneda de prestigio, regulen el ‘juego’ en los campos científicos?<sup>(7)</sup>

Estamos, por tanto, en presencia de otros cuestionamientos que tienen que ver con la propia naturaleza del conocimiento, su producción, acceso y circulación. Estas problematizaciones están enmarcadas en discusiones sobre las influencias del capitalismo neoliberal en las universidades y los sistemas de ciencia, lo que demuestra la tensión latente entre, por una parte, las prácticas y esfuerzos de socialización del conocimiento, a partir de su concepción como bien público y, por otra, la concepción del conocimiento como *commodity*, a merced de las fuerzas del mercado, que empujan hacia la privatización y mercantilización de los conocimientos.

## Métodos

### Selección del *corpus* para el análisis

Teniendo en cuenta que el objetivo del artículo es ejemplificar el uso de *Voyant Tools* para explorar de manera interactiva los textos científicos, el *corpus* está conformado por documentos que cuestionan y problematizan el acceso abierto y la ciencia abierta, con el que ya tengo cierta familiarización, aunque no haya leído a profundidad previamente todos los documentos que lo conforman. De esa manera, puedo evaluar la capacidad de *Voyant Tools* para contribuir con los dos tipos de interrogantes de *Sinclair* e *Rockell*, mencionadas en la sección anterior: ¿servirá para descubrir nuevas perspectivas sobre textos ya conocidos?, ¿será útil para identificar cuáles textos aún no conocidos debería leer porque abordan temas de interés para mi investigación?

El *corpus* está conformado por documentos recuperados a través de búsquedas en las bases de datos *Web of Science*, *Scopus* y *Library and Information Science Abstract* (LISA), bajo las palabras *open access*, *critical studies*, *critique* en los campos título, resumen y palabras clave, así como a través de las referencias bibliográficas de los documentos recuperados. Aunque *Voyant Tools* permite trabajar con textos en diferentes lenguas, decidí trabajar apenas con textos en inglés, que eran mayoría en los resultados, para tener un vocabulario uniforme en esta primera exploración.

### Preparación del *corpus*

El *corpus* quedó conformado por 33 documentos en formato PDF ([Anexo](#)). De ellos, 17 artículos, 13 capítulos de libros, 2 trabajos de eventos y 1 *preprint*, publicados entre los años 2006 y 2020. Fueron renombrados con el primer apellido del autor o autora y el año de publicación para una mejor identificación.

### Cargando el *corpus*

Aunque *Voyant Tools* posee una versión de escritorio, decidí utilizar la interface web, versión 2.4. <https://voyant-tools.org>. La interfaz web permite adicionar los documentos a través de tres alternativas: adicionar directamente el texto en la ventana de inicio (texto plano, HTML, XML) de la interfaz web, insertar el *url* de las páginas web o “subir” (*upload*) los archivos almacenados en la computadora del usuario (en PDF, MS Word, RTF). Los archivos también pueden ser subidos compactados como ZIP. En este caso, utilicé la opción de subir los 33 documentos, seleccionándolos todos de una vez.

Una vez cargados los documentos, procedí a modificar la lista de palabras comunes (*stopwords*). *Voyant Tools* ya trae una lista para el inglés que puede ser personalizada. En este caso excluí, por ejemplo, *http*, *https*, *doi.org*, et al, entre otras que no tenían un significado relevante para los textos. También, utilizando la opción de Modificar documentos, reordené los archivos dentro de la interfaz para que aparecieran ordenados cronológicamente según fecha de



publicación. Este paso es importante si se quiere realizar un análisis diacrónico de los temas.

### Interacción con la herramienta

Una vez cargados los documentos, *Voyant Tools* presenta varios análisis y visualizaciones distribuidos por defecto en una interfaz formada por cinco módulos fundamentales (*Skin*): Cirrus, Lector, Tendencias, Resumen y Contextos. Aunque estas son las opciones por defecto, *Voyant Tools* tiene una caja de herramientas muy amplia que puede ser accedida desde un menú en la esquina derecha de todos estos módulos. Para este artículo se utilizaron las cinco herramientas básicas y, además, Enlaces, Términos, Mandala, gráfico de flujo y ScatterPlot.

Más detalles sobre la interface y sus herramientas pueden ser consultados en la documentación de *Voyant Tools* (<https://voyant-tools.org/docs/#!/guide>) y en el tutorial de *Gutiérrez de la Torre*.<sup>(21)</sup>

## Resultados

A partir de la interfaz por defecto y sus herramientas Resumen (1), Cirrus (2), Tendencias (3), Lector (4) y Contextos (5) podemos tener una primera idea de las características del *corpus* (Fig. 3). La numeración y la secuencia de los módulos se corresponden con el movimiento de análisis que seguí para exploración del *corpus*, una secuencia de “manecillas de reloj”, comenzando por la parte inferior izquierda.

En la parte inferior izquierda (destacado como 1), el módulo Resumen (*Summary*) nos ofrece las características cuantitativas del *corpus*; por ejemplo, cantidad de palabras, extensión de los documentos, documentos más extensos y menos extensos, palabras más frecuentes y palabras diferentes con respecto al vocabulario del *corpus*. Esta visión panorámica es posible para todo el *corpus*, pero también, aunque con menos detalles, para cada documento individualmente, si seleccionamos la opción Documento. En este caso, fue

posible observar las palabras más frecuentes en el *corpus* y la cantidad de ocurrencias (frecuencia absoluta): *open* (2783); *access* (1973); *science* (1315); *research* (1067); *oa* (700).

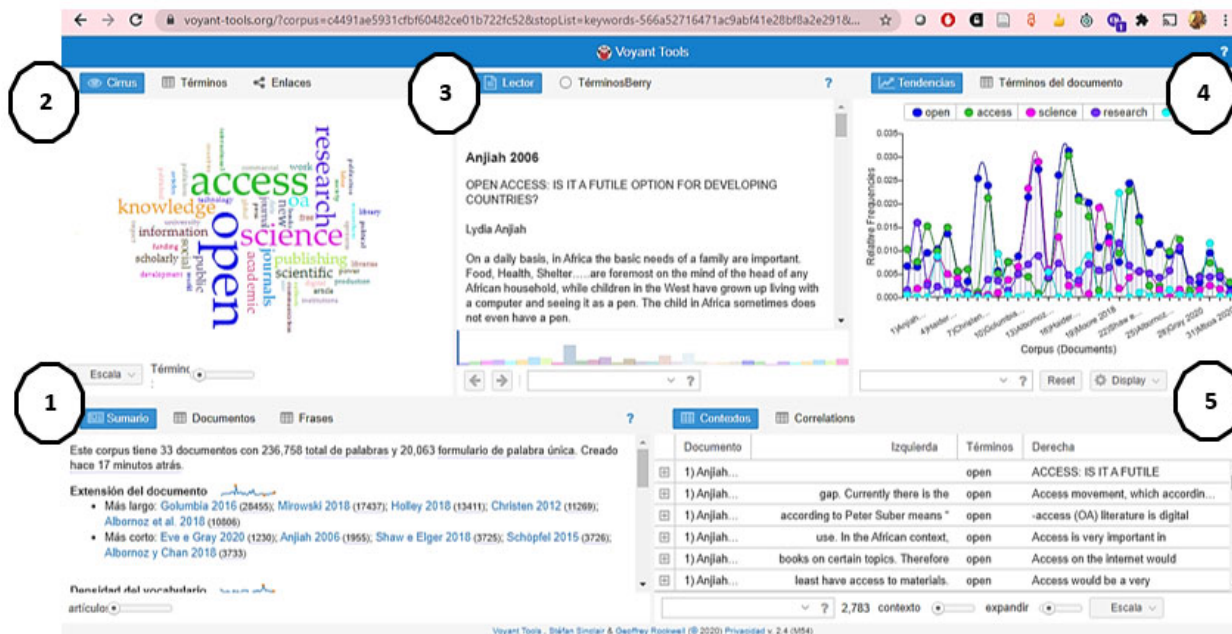


Fig. 3 - Interface de *Voyant Tools* con el corpus cargado para análisis.

No obstante, esto puede observarse con mayor detalle en el módulo 2, tanto por medio de una nube de palabras (aquí llamado Cirrus), donde aparecen representados los términos en diferentes tamaños según su frecuencia de aparición en el *corpus*, como también en formato de lista de términos con sus frecuencias absolutas (Fig. 4). Esta herramienta también muestra la frecuencia relativa de cada palabra por documento en un gráfico de tendencia que permite, si fuera de interés para el analista, clicar en el gráfico para observar con más detalle en que documentos se producen estas ocurrencias. La frecuencia relativa es una medida que tiene en cuenta la frecuencia de cada palabra con respecto a la cantidad de palabras totales de cada documento. Es posible observar la aparición de otras palabras asociadas al campo semántico de acceso abierto y ciencia abierta como *knowledge*, *journals*, *scientific*, *publication*.

	Términos	Contar ↓	Tendencia
1	open	2783	
2	access	1973	
3	science	1315	
4	research	1067	
5	oa	700	
6	knowledge	658	
7	journals	573	
8	scientific	546	
9	publishing	538	

Fig. 4 - Palabras más frecuentes en el *corpus* y su ocurrencia relativa en el *corpus* (tendencia).

El módulo Lector, numerado como 3 en la figura 4, es una herramienta de lectura del texto. Por defecto aparece el primer documento del *corpus*, que, en este caso, como el *corpus* está ordenado por año de publicación, se trata del documento más antiguo ([anexo: Anjiah, 2006](#)). Es uno de los primeros cuestionamientos sobre el acceso abierto, realizado por la documentalista *Lydia Anjiah*, de Tanzania. La autora cuestionaba la relevancia del acceso abierto cuando aspectos básicos como alimentos, alfabetización, electricidad, conexión a internet, disponibilidad de computadoras y bibliotecas, entre otros, aún no se habían resuelto para gran parte de la población en países del Sur, específicamente en África.

Otra herramienta interesante para explorar el *corpus* se encuentra en el módulo de la parte superior derecha (indicado con el número 4), que presenta un gráfico Tendencias, el cual representa la frecuencia relativa de los términos más frecuentes en el *corpus*. Si en el primer módulo o en la herramienta Lector seleccionamos palabras diferentes para analizar, podremos ver cómo ese módulo cambia interactivamente para mostrarnos una visualización que refleja apenas la o las palabras seleccionadas. En la figura 3 en este módulo se visualizan las tendencias de las cinco primeras palabras más frecuentes en el *corpus* estudiado.

El módulo de la derecha abajo (indicado con el número 5), llamado Contexto, precisamente nos permite ver las palabras en contexto, presentando la frase

anterior y posterior de la palabra seleccionada. En la figura 3 se observan la palabra *open* en los contextos de aparición en el texto de Anjiah (2006) comentado antes.

En el módulo 1, la herramienta Enlaces (*Collocates Graph*) resulta interesante para revelar relaciones temáticas dentro del *corpus*, pues se trata de un gráfico de red que representa la relación entre un término y aquellos con los que guarda relaciones de proximidad en el *corpus*. En la figura 5, panel A (derecha) es posible ver la red de términos asociados a las tres palabras más frecuentes: *open*, *access* y *science* (en azul). La construcción de la visualización fue interactiva, o sea, en la medida en que se hace doble clic en un término, sus principales relaciones van emergiendo. Así, en el panel B (derecha) es posible identificar la emergencia de nuevos conglomerados: *open-access-science-policies-source*; *access-policies*; *science-open-access-technology-policies*; *journals-publishing-open-access-high*; *high-impact-quality-open-access*; *policies-discourse*; *open-access-source-code-free-software*.

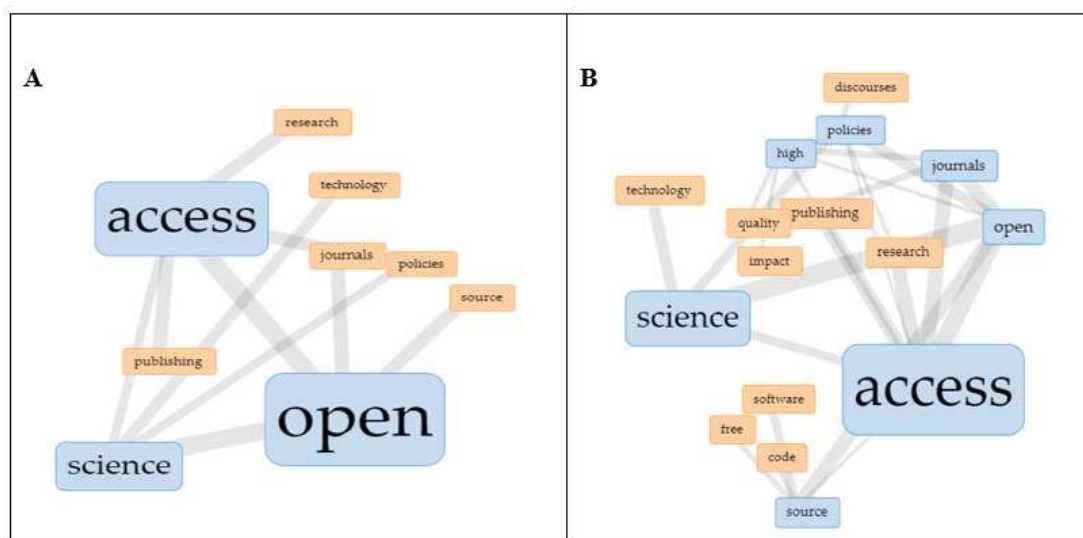
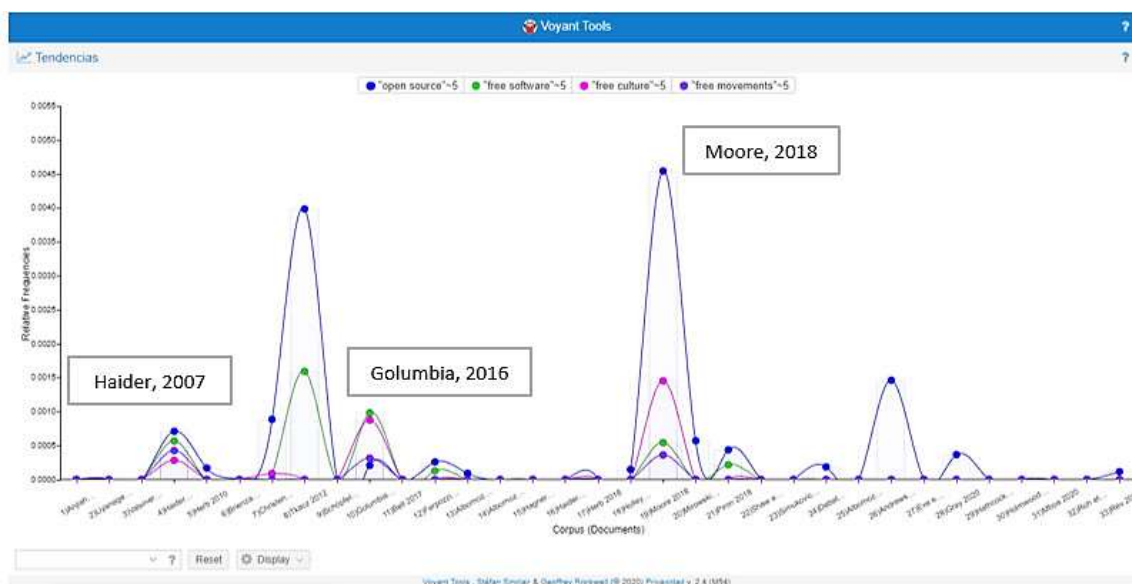


Fig. 5 - Red de términos más frecuentes y palabras comúnmente relacionadas.

Entre los términos que emergieron en esta visualización están *source-code-free-software*. Estos no aparecen entre los primeros resultados de la lista de términos según frecuencia y, sin embargo, gracias a esta herramienta se visibilizaron, invitando a una exploración más profunda. Por ejemplo, ¿Cuál es

la relación entre el acceso abierto y los movimientos de “cultura libre”, como el software libre y el código abierto? ¿Qué críticas o problematizaciones se discuten en el *corpus* acerca de esta relación? Para responder a estas indagaciones, los enfoques son múltiples, pero uno de ellos podría ser identificar cuales autores están abordando estos temas y descubrir que discuten. Para esto, el procedimiento adoptado fue hacer una búsqueda de las palabras compuestas (*free software, open source, free culture, free movements*), directamente en la caja de búsqueda del módulo Términos.

Teniendo en cuenta que la gramática inglesa permitiría utilizar estas palabras no necesariamente en secuencia, utilicé la sintaxis que permite recuperar los términos en una proximidad de hasta cinco palabras (ejemplo “free software” ~5). La selección de estas cuatro palabras transforma automáticamente la visualización del módulo 3 Tendencias como se observa en la figura 6. En el eje X están los documentos, identificados por el apellido del autor y año de publicación, y en el eje Y la frecuencia relativa de las palabras en el *corpus*.



**Fig. 6 -** Módulo Tendencias. Representación de las frecuencias relativas de las palabras seleccionadas (*open source, free software, free culture, free movements*) por documento.

En este caso, podemos identificar que en los documentos *Moore* (2018), *Golumbia* (2016) y *Haider* (2007) aparecen los cuatro términos, pero es en

Moore (2018) donde los términos aparecen con mayor frecuencia relativa. Tkacz (2012), por otra parte, también tiene una alta frecuencia relativa de los términos *open source* y *free software*. Estas indicaciones nos dan buenas pistas de por donde comenzar para conocer cuáles son las relaciones entre acceso abierto y estos otros movimientos de apertura y cuáles son las problematizaciones que se plantean.

Interactuando con esa visualización podemos obtener más detalles. Por ejemplo, al clicar en la línea de tendencia de la palabra *open source*, en el documento Moore (2018), automáticamente esta palabra va a aparecer en el módulo Contextos, rodeada, a derecha y a izquierda por frases de contexto. En este caso, me ha llamado la atención una frase que dice “*to see a link entre open source and neoliberalism*”. La captura de pantalla de este módulo que se observa en la figura 7 muestra esta frase y el párrafo donde aparece. Moore (2018) analiza la existencia de vínculos entre la filosofía del movimiento de código abierto y el neoliberalismo, por su alineación al enfoque de mercado y al modelo descentralizado de gobierno desarrollado por teóricos iniciales del neoliberalismo como *Popper* y *Hayak*.

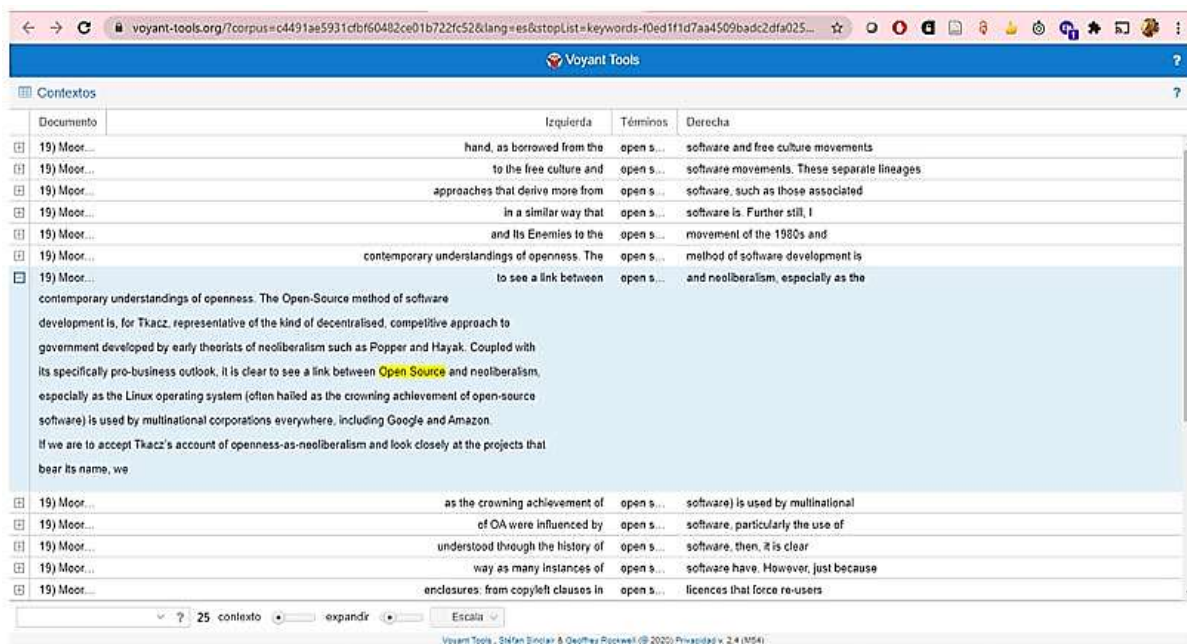


Fig. 7 - Contextos de “open source” en el documento Moore (2018). Relación con neoliberalismo.

Explorando el contexto en que ocurre “*open source*” aparece la discusión sobre los diferentes linajes del acceso abierto y como algunos de ellos vienen de esa rama de los movimientos de código abierto que los autores analizan como vinculados con el neoliberalismo. Entre las manifestaciones de esta relación está el énfasis en la idea de que la diseminación del conocimiento requiere libertades irrestrictas de uso y reúso de los productos de conocimiento, incluyendo uso comercial, y promoviendo la utilización licencias *Creative Commons* CC-BY. Entre los adeptos de esta vertiente están las editoras de revistas de acceso abierto comerciales que hoy tienen el cobro de cargos por procesamiento de artículos (APC), como su fuente principal de lucros.

Si regresamos a la figura 5, vemos que otras relaciones interesantes emergieron, por ejemplo, entre factor de impacto y acceso abierto (*High-impact-quality-open-access*). De la revisión de literatura sabemos que una de las críticas más importantes actualmente es la orientación mercantil de una parte de las iniciativas de acceso abierto y la vinculación con el factor de impacto. Para continuar explorando como estos temas son abordados en el corpus, en la caja de búsqueda de la herramienta Términos busqué por *apc\** (con el comodín de truncado para recuperar también *apcs*), *gold*, “*impact factor*” y *prestig\** (*prestige*, *prestigious*).

A partir de esa búsqueda, verifiqué varias visualizaciones. Una visualización interesante se denomina Mandala. Muestra las relaciones entre los términos seleccionados y los documentos. Cada término actúa como un imán que atrae documentos hacia él en función de la frecuencia relativa del término en el corpus. Si colocamos el cursor encima de un término se destacan los documentos que lo contienen. En la figura 8A se muestra la Mandala y en la Fig. 8 B una ampliación de esta representación, mostrando la relación entre el término “*apc\**” y los documentos que lo contienen. Una vez más la visualización nos da indicios de cuáles documentos explorar para profundizar en estos temas.

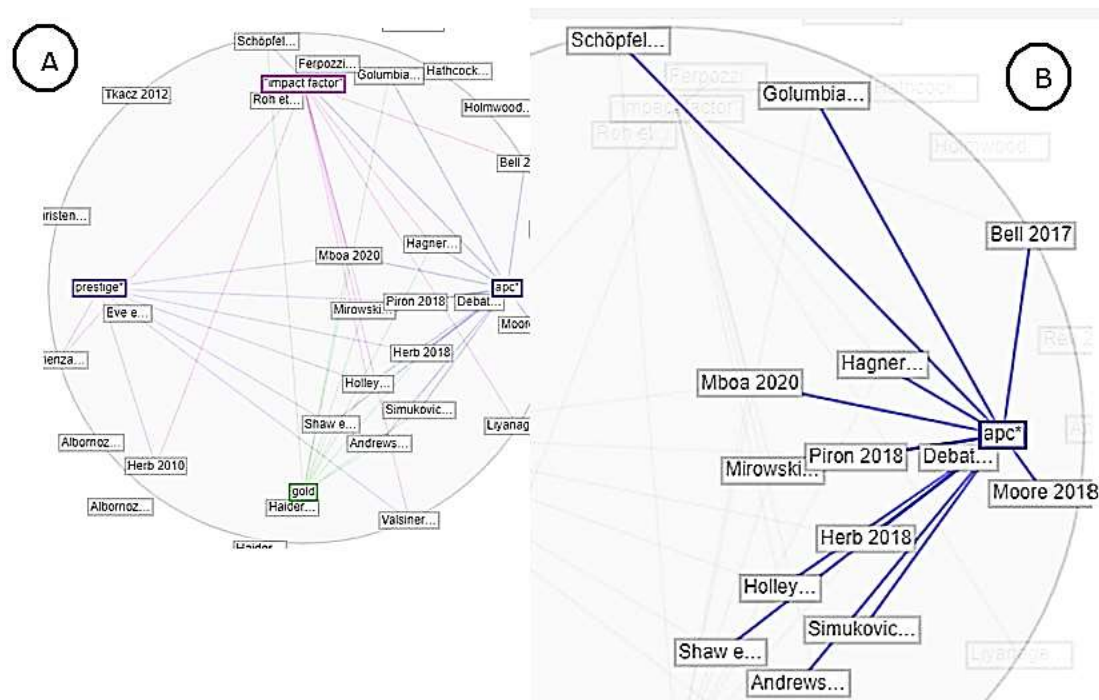


Fig. 8 - A: Se muestra la Mandala. B: Relaciones entre los términos seleccionados y los documentos que lo contienen.

Una de las discusiones más pertinentes sobre este tema es que la adopción generalizada del modelo basado en APC por parte de las editoriales comerciales, el aumento sostenido de los valores cobrados (un promedio de 1500 USD por artículo) y la demanda creciente de publicación en estas revistas a pesar de estos aumentos,<sup>(34)</sup> evidencia la imbricación entre el prestigio percibido por la publicación en determinados canales, el uso del indicador factor de impacto como indicador de ese prestigio y los valores de APC. O sea, revistas como mayor factor de impacto son percibidas como de mayor prestigio, por lo que son más atractivas, sin importar el precio que se cobre por publicar en ellas. Un alto precio también refuerza la percepción de prestigio como capital simbólico que se transfiere a quienes puedan asumir ese costo.

La relación entre los términos también puede verse en el gráfico de red (Fig. 9) a partir de los términos *gold*, *apc\**, *prestig\** e “*impact factor*”. Es interesante la emergencia de la palabra exclusión (*exclusion*), vinculada a prestigio. Semánticamente, prestigio también está asociado a distinción, diferencia,



exclusividad. Los términos y sus asociaciones evidencian problematizaciones con relación al factor de impacto y el cobro de APC.

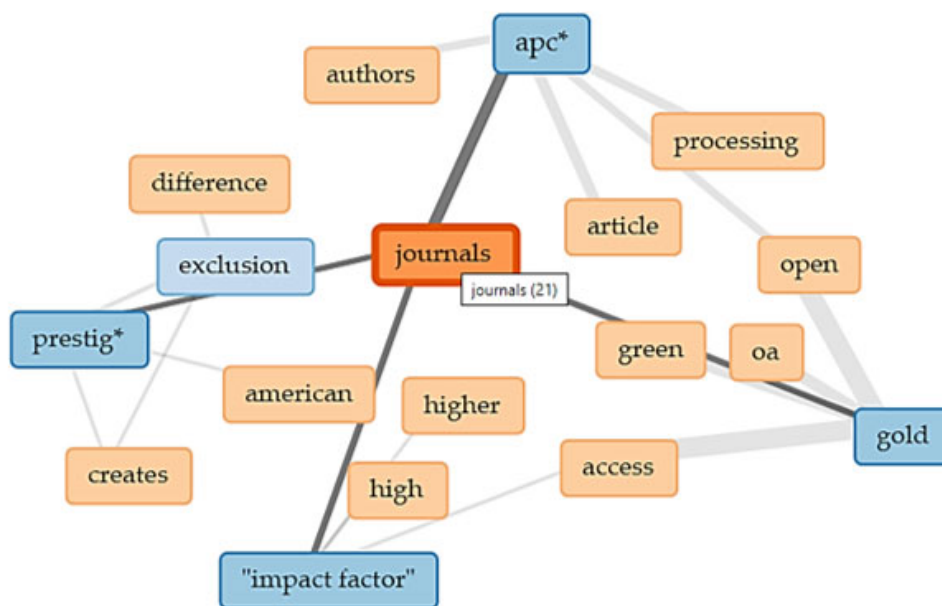


Fig. 9 - Red de palabras asociadas a los términos *gold*, *apc\**, *prestig\** e “*impact factor*”.

Explorando el término “*exclusion*” en la herramienta de Tendencias identifiqué que las mayores frecuencias relativas se dan en los documentos de *Herb* (2018) y *Mboa* (2020). Para profundizar en la manera en que los autores problematizan este concepto, podemos utilizar la herramienta Contextos. La figura 10 muestra frases del capítulo de *Herb* (2018), titulado *Open Access and Symbolic Gift Giving*, alrededor del término *exclusion*.

Voyant Tools			
Contextos			
Documento	Izquierda	Términos	Derecha
17) Herb ...	an instrument that creates exclusivity,	exclusion	, distinction and prestige. These functions
17) Herb ...	impact but high-priced journals. Surprisingly, this transformation of open access is not the subject of any noteworthy discussion in specialist or journalistic publications, but instead the ideals of the digital commons of knowledge still prevail in these venues. Even so open access is increasingly becoming an instrument that creates exclusivity, <b>exclusion</b> , distinction and prestige. These functions, however, are obscured by symbolic gift giving strategies and presented as altruistically staged, so that in the discourse of the open access community and in media reporting on open access, the both ephemeristic and largely obsolete prosocial story-telling of open access dominates. The paper		
17) Herb ...	often reached by means of	exclusion	. But how could a project
17) Herb ...	Conclusion Social Fields and Distinction	exclusion	Symbolic Goods and Symbolic Gift
17) Herb ...	morph into a program deeply	exclusi...	in character? A glance at
17) Herb ...	for distinctiveness and power" (14).	exclusi...	stated before, distinction and distinctiveness
	raise (...) reputations, build exclusive groups and exclude others. No matter how open a network or infrastructure [addition by the author: or principle] might be, they are devised, shaped and used by individuals. Consequently, their utilization is subject to human interests and necessities like networks constructed for distinctiveness and power" (14).		
	<b>ExclusionAs</b> stated before, distinction and distinctiveness are often reached by means of exclusion. But how could a project launched with the expressly stated goal of uniting humanity morph into a program deeply exclusionary in character? A glance at the history of open access helps to understand this unexpected turnaround, which is		

Fig. 10 - Contextos en que ocurre la palabra “*exclusion*” en el documento de *Herb* (2018).

El autor argumenta que el auge de las revistas de acceso abierto comerciales basadas en APC está transformando el acceso abierto en un instrumento de distinción, prestigio y exclusividad. La exclusividad se da también por medio de la exclusión de aquellos que no pueden pagar (por ejemplo, los investigadores de las naciones del Sur). Incluso, alerta que los científicos han comenzado a elegir a que universidades o instituciones van a trabajar en función de la capacidad de estas de cubrir los APC que cobran las revistas de alto factor de impacto. Para *Herb*, lo que parecería ser un comportamiento altruista (por parte del investigador y de la institución que cubre los APC) se transforma en una estrategia de poder donde el acceso abierto a la producción institucional es un obsequio simbólico que refuerza su posición de poder, distinguiéndose de sus competidores y atrayendo talentos.

*Voyant Tools* permite también algunos análisis y visualizaciones a nivel de documento individual. En este caso, podemos utilizar esa posibilidad para profundizar en las características del texto *Mboa* (2020), un capítulo titulado *Epistemic Alienation in African Scholarly Communications: Open Access as a Pharmakon*. La herramienta *Cirrus* ofrece un panorama general de las palabras del documento (Fig. 11). Podemos ver destacadas palabras como *african*, *knowledge*, *epistemic*, *alienation*, *south* y *coloniality*. Estas palabras sugieren que su trabajo aborda aspectos críticos relacionados con el papel del acceso abierto en la producción de conocimiento de países del Sur, con destaque para el caso de los países africanos.



Fig. 11 - Nube de palabras conformada por el vocabulario principal de *Mboa* (2020).

Para confirmar podemos acudir nuevamente a la herramienta Contextos o a la herramienta Lector. Al hacerlo, descubrimos que, reformulando la metáfora de *pharmakon* como aquello que puede ser a la misma vez cura y veneno, Mboa destaca que para los países del Sur el acceso abierto puede ser considerado una especie de *pharmakon*, con beneficios, por una parte, pero también con aspectos negativos, haciendo énfasis en la orientación mercantilista de declaraciones, políticas y modelos de publicación en acceso abierto que ahondan las brechas entre Norte y Sur, creando nuevas formas de colonialismo y exclusión.

Una exploración en el *corpus* por las palabras *colonialism*, *neocolonialism* (y sus formas derivadas) y *coloniality* evidencia que muchos de los trabajos incluidos en este estudio abordan estos conceptos. La figura 12 es una variante visualmente interesante del gráfico de Tendencia, esta vez representando las frecuencias relativas de los términos en forma de flujo. Como los trabajos están organizados cronológicamente por año de publicación, puede observarse que los términos asociados a “colonial\*” aparecen desde el primer trabajo en 2006 con una presencia casi constante en todo el periodo estudiado.

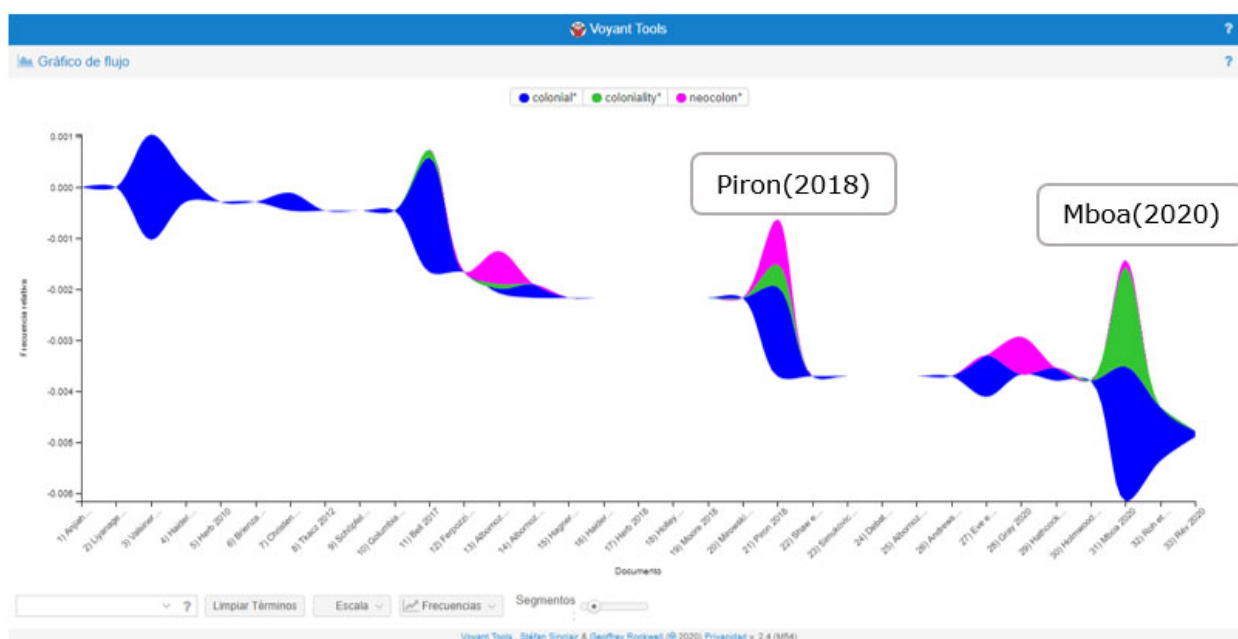


Fig. 12 - Flujo de frecuencias relativas de los términos “colonial\*”, “neocolonial\*” y “coloniality”.

En varios trabajos las problematizaciones tienen como base teorías feministas, poscoloniales y decoloniales que denuncian como el conocimiento científico occidental tiene un carácter excluyente y cómplice de las prácticas colonialistas, que ha contribuido con el epistemicidio, o sea, con la destrucción y discriminación de otros conocimientos y prácticas culturales.<sup>(35)</sup> Se critica también su contribución a la perpetuación de las relaciones de subordinación/exclusión de los pueblos que un día fueron colonizados, a través de estructuras de poder que *Aníbal Quijano* y *Edgardo Lander* han denominado colonialidad del poder y colonialidad del saber, respectivamente.<sup>(36)</sup>

*Piron* (2018) y *Mboa* (2020), por ejemplo, son trabajos que incluyen los tres términos, lo que evidencia también una discusión enmarcada en las teorías decoloniales. Por ejemplo, una de las críticas en este sentido con relación al acceso abierto es que un mayor acceso a las revistas científicas en inglés producida por los países del Norte, que son los canales que continúan garantizando mayor prestigio y reconocimiento, podría perpetuar la subordinación epistémica (o alienación epistémica, en palabras de *Mboa*) de los académicos del Sur y la destrucción de otras formas diversas de epistemes, incluyendo las lenguas nacionales.

Finalmente, para tener una idea panorámica de la relación entre temas y documentos, podemos usar la herramienta *ScatterPlot* que crea *clusters* o conglomerados de palabras basados en semejanza de los documentos, análisis de correspondencia y análisis de componentes principales. Para ejemplificar, en la caja de búsqueda de Términos introduje algunos de los utilizados en los análisis anteriores: *free software*, *open source*, *apc\**, *impact factor*, *colonial\**. La figura 13 es una representación creada con *ScatterPlot* de las semejanzas y diferencias entre los vocabularios, a partir de un análisis de correspondencia entre las palabras y los documentos.



Fig. 13 - ScatterPlot. Semejanza (y diferencias) de vocabulario de los documentos del corpus.

A partir del mapa es posible identificar, por lo menos, tres conglomerados. Los trabajos del cuadrante superior derecho, marcados con los rombos color magenta, están relacionados con las palabras *free software* y *open source*. En estos documentos se identifican diversas problematizaciones, por ejemplo, las situaciones de apropiación y comercialización de conocimientos abiertos, sin compensación para sus productores (explotación cognitiva); la orientación hacia el mercado de las estrategias y filosofías que permean el acceso abierto y la ciencia abierta; el predominio de los discursos enfocados en aceleración capitalista, maximización de lucros y de competitividad en las políticas y estrategias de ciencia abierta, entre otras.

En el cuadrante izquierdo (rombos azules), los documentos se asocian en torno a \*apc. En este caso, los trabajos discuten fundamentalmente, las transformaciones y consecuencias del modelo de acceso abierto adoptado por revistas comerciales, incluyendo las consecuencias de su generalización por medio de políticas que pretenden mundializarse como el Plan S.<sup>(37)</sup>

Los documentos en rombos verdes, asociados a las palabras \**impact factor* y *colonial*\* problematizan, por ejemplo, el abuso del indicador factor de impacto como supuesta medida de calidad y prestigio de las revistas y de sus autores, y

las distorsiones de los sistemas de evaluación, que crean cambios en las prácticas de publicación de los autores e, incluso, promueven la emergencia de revistas de acceso abierto “depredadoras”. Otros trabajos discuten como la distorsión en las prácticas de publicación de los autores causada por la presión de publicar en revistas con factor de impacto puede llevar, sobre todo a los del Sur, a abandonar sus lenguas, sus revistas nacionales, sus agendas de investigación para encajar en el tipo de ciencia que se publica en circuitos de publicación controlados por países del Norte, en clara manifestación de colonialismo y epistemicidio. Las revistas de acceso abierto comerciales que cobran APC refuerzan estas distorsiones.

Aunque es posible identificar énfasis distintos en los tres conglomerados, evidentemente, hay hilos comunes entre ellos: la orientación mercantilista que ha adoptado una parte creciente de las iniciativas de acceso abierto y ciencia abierta, evidenciada en la expansión del modelo de revistas de acceso abierto comerciales basadas en APC y la presión por la adopción de licencias Creative Commons CC-BY que permiten usos comerciales, configurándose como la norma para los modelos de acceso abierto; la extensa adquisición de infraestructuras de ciencia abierta por parte de los oligopolios editoriales, dominando todo el circuito de la ciencia,<sup>(38)</sup> el uso extendido del factor de impacto como medida de calidad y prestigio, asociada a índices y revistas comerciales y las consecuencias negativas de estos desarrollos para el Sur Global en sus aspiraciones de equidad y justicia social y cognitiva.

## Conclusiones

En este artículo se muestra la utilización de *Voyant Tools* para analizar y visualizar textos científicos, utilizando como objeto de estudio un *corpus* conformado por 33 documentos que problematizan diferentes aristas relacionadas con los movimientos de acceso abierto y ciencia abierta. *Voyant Tools* tiene como ventajas ser gratuito, trabajar en entorno web con diferentes formas de entrada de textos, múltiples formatos de archivo y diferentes lenguajes (para los textos y la interface). Además, reúne de manera compacta

numerosos tipos de herramientas analíticas (análisis de temas, de tendencias, de correspondencia, sintácticos, lexicales, entre otros) y de visualización (exploración, navegación, *clusters*, clasificaciones, categorías), así como representaciones gráficas variadas (nube, líneas, columnas, área). Fue posible verificar que estas herramientas sirven para diversos propósitos.

Particularmente, en este estudio permitieron explorar el *corpus* desde perspectivas distintas y complementarias; mostrar los contenidos de los documentos (por ejemplo, palabras y frases en contexto) y sus características cuantitativas, así como semejanzas y diferencias entre los vocabularios de los documentos, y, por tanto, sus temas. Uno de los aspectos más interesantes fue la interactividad de las herramientas, pues la selección o introducción (por medio de las cajas de búsqueda) de una palabra en uno de los módulos, por ejemplo, automáticamente transforma las visualizaciones de todos los otros módulos, y crea nuevas provocaciones.

A través de los ejemplos aquí presentados espero haber mostrado algunas de las potencialidades de *Voyant Tools* para favorecer la visualización e interpretación de los documentos. Se trata de un ciclo hermenéutico, pues en la medida en que exploramos el *corpus* con las herramientas de análisis y visualización, descubrimos nuevas pistas que, como un rastro de migajas de pan, nos convidan a otras exploraciones, visualizaciones y hallazgos. Fue constatada su utilidad para profundizar en la comprensión de textos ya conocidos; sin embargo, el camino de descubrimientos y exploraciones mostrado en este artículo también evidencia las posibilidades de la herramienta para analizar textos que no conocemos previamente.

En relación con la temática explorada, su multidimensionalidad hace que sean diversas las perspectivas y los temas abordados por los autores de los documentos incluidos en este estudio. En este caso, la herramienta permitió la identificación de términos y temas indicativos de críticas y problematizaciones sobre el acceso abierto y la ciencia abierta, apuntando directamente a autores y partes de textos y sugiriendo conglomerados de documentos reunidos por su

semejanzas lingüísticas y semánticas. A partir de aquí, depende del investigador una lectura más direccionada y detenida que permita la comprensión cabal de los temas y relaciones identificados.

Entre los aspectos relacionados con críticas y problematizaciones sobre el acceso abierto, se encuentran, por ejemplo, los linajes del acceso abierto vinculados con los movimientos de cultura libre (específicamente con el movimiento de código abierto) y la vinculación de este último con ideologías neoliberales, aspecto que también está presente en políticas, discursos y estrategias relacionadas con el acceso abierto y la ciencia abierta. Fue observada también la relación entre el factor de impacto como indicador de prestigio y las revistas de acceso abierto que cobran APC, lo que se convierte en nuevos instrumentos de prestigio y exclusión. Identificamos también que cuando esta relación malsana entre factor de impacto y revistas de acceso abierto comerciales se analiza en el contexto de la producción y circulación de conocimientos global, donde el valor de los conocimientos depende del medio y la lengua en que han sido publicados, entonces emergen los cuestionamientos sobre la colonialidad del poder y del saber, como patrones de relaciones de inferioridad y subordinación epistémica que movimientos de abertura acríticos podrían perpetuar.

## Referencias bibliográficas

1. Lima JLO, Álvares L. Organização e representação da informação e do conhecimento. En: Álvares L, editor. Organização da informação e do conhecimento: conceitos, subsídios interdisciplinares e aplicações. São Paulo: B4 Editores; 2012. p. 21-48.
2. Alvarenga L. Representação do conhecimento na perspectiva da. Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação. 2003;(15):23.
3. Vargas-Quesada B, Moya-Anegón F. Visualizing the Structure of Science. Berlin: Springer-Verlag; 2007.
4. Sinclair S, Rockwell G. Text Analysis and Visualization: Making Meaning Count. En: Schreibman S, Siemens R, Unsworth J, editores. A New Companion to Digital Humanities. West Sussex: John Wiley & Sons, Ltd; 2016. p. 274-90.



5. Sinclair S, Rockwell G. About - Voyant Tools Help. 2020 [acceso: 27/12/2020]. Disponible en: <https://voyant-tools.org/docs/#!/guide/about>
6. Tennant JP, Waldner F, Jacques DC, Masuzzo P, Collister LB, Hartgerink ChrisHJ. The academic, economic and societal impacts of Open Access: an evidence-based review. *F1000Res*. 2016 [acceso: 22/12/2020];5:632. Disponible en: <https://f1000research.com/articles/5-632/v1>
7. Sánchez-Tarragó N. Ciência aberta e acesso aberto para o Sul: perspectivas críticas e desafios. En: Moreira LA, Souza JA, Tanus GF, editores. *Informação na Sociedade Contemporânea*. Florianópolis: Rocha Gráfica e Editora (Selo Nyota); 2020.
8. González Aguilar A, Pinto AL, Semeler AR, Soares APA. *Visualização de dados, informação e conhecimento*. Florianópolis: Editora UFSC; 2017.
9. Ward M, Grinstein GG, Keim D. *Interactive data visualization: foundations, techniques, and applications*. Second edition. Boca Raton: CRC Press, Taylor & Francis Group; 2015. p. 558.
10. Torres Ponjuán D, Ponjuán Dante G. Influencias ergonómicas en la visualización de la información. *Informação & Informação*. 2010 [acceso: 27/12/2020];15(1):17-36. Disponible en: <http://www.uel.br/revistas/uel/index.php/informacao/article/view/3895>
11. Caballero-Rivero A. Mapeamento exploratório dos estudos sobre Visual Analytics: principais conceitos, metodologias e ferramentas. En: Moreira LA, Souza JA, Tanus GF, editores. *Informação na Sociedade Contemporânea*. Florianópolis: Rocha Gráfica e Editora (Selo Nyota); 2020. p. 51-74.
12. Börner K, Chen C, Boyack KW. Visualizing knowledge domains. *Annual Review of Information Science and Technology*. 2003 [acceso: 27/12/2020];37(1):179-255. Disponible en: <https://asistdl.onlinelibrary.wiley.com/doi/abs/10.1002/aris.1440370106>
13. Jockers ML, Underwood T. Text-Mining the Humanities. En: Schreibman S, Siemens R, Unsworth J, editores. *A New Companion to Digital Humanities*. West Sussex: John Wiley & Sons, Ltd; 2016. p. 291-306.
14. Cai Y, Sun J-T. Text Mining. En: Liu L, Özsu MT, editores. *Encyclopedia of Database Systems* [Internet]. Boston, MA: Springer EE.UU.; 2009. p. 3061-5. DOI: [https://doi.org/10.1007/978-0-387-39940-9\\_418](https://doi.org/10.1007/978-0-387-39940-9_418)
15. Gálvez C. Minería de textos: la nueva generación de análisis de literatura científica en biología molecular y genómica. *Enc Bibli: R Eletr Bibliotecon Ci Inf*. 2008 [acceso: 27/12/2020];13(25):1-14. Disponible en: <https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2008v13n25p1>

16. Cao N, Cui W. Introduction to Text Visualization. Paris: Atlantis Press; 2016 [acceso: 27/12/2020]. Disponible en: <http://link.springer.com/10.2991/978-94-6239-186-4>
17. Kucher K, Kerren A. Text visualization techniques: Taxonomy, visual survey, and community insights. En: 2015 IEEE Pacific Visualization Symposium (PacificVis). 2015. p. 117-21.
18. Kaur A, Chopra D. Comparison of Text Mining Tools. En IEEE; 2016 [acceso: 27/12/2020]. Disponible en: <https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=7784950>
19. Hofmann M, Chisholm A, editores. Text Mining and Visualization: Case Studies Using Open-Source Tools [Internet]. Boca Ratón (FL): Chapman and Hall/CRC; 2016 [acceso: 28/12/2020]. Disponible en: <https://www.taylorfrancis.com/books/9781482237580>
20. Rockwell G, Sinclair S. Hermenéutica: computer-assisted interpretation in the humanities. Cambridge, Massachusetts ; London, England: The MIT Press; 2016:246.
21. Gutiérrez de la Torre S. Análisis de *corpus* con Voyant Tools. Programming Historian en español. 2019 [acceso: 11/03/2020]. Disponible en: <https://programminghistorian.org/es/lecciones/analisis-voyant-tools>
22. Nascimento FA, Massoni LFH, Shirakava RDS, Pinho FA, Martínez-Ávila D. Automeaçã e autoclassificação das homossexualidades masculinas e sexualidades alternativas no Brasil. IB. 2020 [acceso: 28/12/2020];34(84):151. Disponible en: <http://rev-ib.unam.mx/ib/index.php/ib/article/view/58209>
23. Hendrigan H. Mixing Digital Humanities and Applied Science Librarianship: Using Voyant Tools to Reveal Word Patterns in Faculty Research. Issues in Science and Technology Librarianship. 2019;(91).2.
24. Cortés Sánchez JD. Mission statements of universities worldwide: Text mining and visualization. intangible capital. 2018 [acceso: 28/12/2020];14(4):584. Disponible en: <http://www.intangiblecapital.org/index.php/ic/article/view/1258>
25. Villada YA. Archival management in Colombia: Lexicometric analysis of the legal framework over two periods, 1913-1989 and 1990-2017. Trilogia. 2018 [acceso: 28/12/2020];10(19):115-32. Disponible en: <https://www.doaj.org>
26. Miller A. Text Mining Digital Humanities Projects: Assessing Content Analysis Capabilities of Voyant Tools. Journal of Web Librarianship. 2018 [acceso: 28/12/2020];12(3):169-97. Disponible en: <https://www.tandfonline.com/doi/full/10.1080/19322909.2018.1479673>

27. Budapest Open Access Initiative. Read the Budapest Open Access Initiative. 2002 [acceso: 09/02/2020]. Disponible en:  
<https://www.budapestopenaccessinitiative.org/read>
28. Sánchez Tarragó N. El movimiento de acceso abierto a la información y las políticas nacionales e institucionales de autoarchivo. ACIMED. 2007 [acceso: 19/12/2020];16(3). Disponible en:  
[http://scielo.sld.cu/scielo.php?script=sci\\_abstract&pid=S1024-94352007000900005&lng=es&nrm=iso&tlng=es](http://scielo.sld.cu/scielo.php?script=sci_abstract&pid=S1024-94352007000900005&lng=es&nrm=iso&tlng=es)
29. Moore S. A genealogy of open access: negotiations between openness and access to research. *Revue française des sciences de l'information et de la communication*. 2017 [acceso: 19/02/2020];(11). Disponible en:  
<https://journals.openedition.org/rfsic/3220>
30. Alperin JP, Fischman G, editores. Hecho en Latinoamérica: acceso abierto, revistas académicas e innovaciones regionales. Buenos Aires: CLACSO; 2015:122.
31. Sánchez-Tarragó N, Fernández-Molina JC, Caballero-Rivero A. Reflexiones en torno al acceso abierto a la información en el contexto cubano: el caso del sector Salud. *Informação & Sociedade: Estudos*. 2012 [acceso: 22/12/2020];22(2). Disponible en: <https://periodicos.ufpb.br/index.php/ies/article/view/10499>
32. Gold and Diamond open access journals landscape. Research Consulting. 2020 [acceso: 28/12/2020]. Disponible en: <https://www.research-consulting.com/an-interactive-look-at-the-gold-and-diamond-journals-landscape/>
33. Halsted CH. Copyright protection and open access. *Am J Clin Nutr*. 2003 [acceso: 16/01/2020];78(5):899-901. Disponible en:  
<https://academic.oup.com/ajcn/article/78/5/899/4677495>
34. Khoo S. Article Processing Charge Hyperinflation and Price Insensitivity: An Open Access Sequel to the Serials Crisis. *LIBER Quarterly*. 2019 [acceso: 31/01/2020];29(1):1-18. Disponible en:  
<http://www.liberquarterly.eu/article/10.18352/lq.10280/>
35. Santos BS. Para além do pensamento abissal: das linhas globais a uma ecologia de saberes. *Novos estudos CEBRAP* [Internet]. noviembre de 2007 [citado 5 de septiembre de 2020];(79):71-94. Disponible en:  
[http://www.scielo.br/scielo.php?script=sci\\_abstract&pid=S0101-33002007000300004&lng=pt&nrm=iso&tlng=pt](http://www.scielo.br/scielo.php?script=sci_abstract&pid=S0101-33002007000300004&lng=pt&nrm=iso&tlng=pt)
36. Lander E, Castro-Gómez S, editores. La colonialidad del saber: eurocentrismo y ciencias sociales: perspectivas latinoamericanas. Buenos Aires: Consejo

Latinoamericano de Ciencias Sociales-CLACSO ; UNESCO, Unidad Regional de Ciencias Sociales y Humanas para América Latina y el Caribe; 2000.

37. Debat H, Babini D. Plan S en América Latina: Una nota de precaución. Peer J Preprints. 2019;11. Disponible en: <https://peerj.com/preprints/27834/>

38. Posada A, Chen G. Inequality in Knowledge Production: The Integration of Academic Infrastructure by Big Publishers. En: Chan L, Mounier P, editores. ELPUB; 2018 [acceso: 09/02/2020]. Disponible en: <https://hal.archives-ouvertes.fr/hal-01816707>

Anexo - Documentos que conforman el *corpus* utilizado en este estudio

1. Albornoz D, Chan L. Power and Inequality in Open Science Discourses Poder e desigualdade nos discursos da Ciência Aberta. Iris - Inform, Mem Technol. 2018);4(1):70-9.

2. Albornoz D, Huang M, Martin IM, Mateus M, Touré AY, Chan L. Framing Power: Tracing Key Discourses in Open Science Policies. 22nd International Conference on Electronic Publishing; 2018. DOI: <https://doi.org/10.4000/proceedings.elpub.2018.23>

3. Albornoz D, Okune A, Chan L. Can Open Scholarly Practices Redress Epistemic Injustice? En: Eve MP, Gray J. (Eds.). Reassembling Scholarly Communications: Histories, Infrastructures, and Global Politics of Open Access. The MIT Press; 2020:65-79. DOI: <https://doi.org/10.7551/mitpress/11885.001.0001>

4. Andrews PCS. The Platformization of Open. En: Eve MP, Gray J. (Eds.). Reassembling Scholarly Communications: Histories, Infrastructures and Global Politics of Open Access. The MIT Press; 2020:265-76. DOI: <https://doi.org/10.7551/mitpress/11885.001.0001>

5. Anjiah L. Open access: Is it a futile option for developing countries? Proceedings from the Coady International Institute: The Open Access Movement and Information for Development; 2006. Disponible en: <https://coady.stfx.ca/wp->

[content/uploads/pdfs/resources/publications/openaccess/anjiah.pdf](https://content/uploads/pdfs/resources/publications/openaccess/anjiah.pdf)

6. Bell K. 'Predatory' Open Access Journals as Parody: Exposing the Limitations of 'Legitimate' Academic Publishing. TripleC: Communication, Capitalism & Critique. Open Access J Glob Sustain Inform Soc. 2017;15(2):651-62.  
<https://doi.org/10.31269/triplec.v15i2.870>
7. Brienza C. Opening the Wrong Gate? The Academic Spring and Scholarly Publishing in the Humanities and Social Sciences. Publish Res Quart. 2012;28(3):159-71. <https://doi.org/10.1007/s12109-012-9272-5>
8. Christen K. Does Information Really Want to be Free? Indigenous Knowledge Systems and the Question of Openness. Internat J Communic. 2012;6:2870-93.
9. Debat H, Babini D. Plan S in Latin America: A precautionary note. Canadian Institute for Studies in Publishing; 2020. DOI: <https://doi.org/10.7287/peerj.preprints.27834v2>
10. Eve MP, Gray J. Reassembling Scholarly Communications: Histories, Infrastructures and Global Politics of Open Access. The MIT Press; 2020:361-3. DOI: <https://doi.org/10.7551/mitpress/11885.001.0001>
11. Ferpozzi H. Public participation and the co-production of open scientific knowledge: What is at stake? Inform Serv Use. 2017;37:451-61. DOI: <https://doi.org/10.3233/ISU-170860>
12. Golumbia D. Marxism and Open Access in the Humanities: Turning Academic Labor against Itself. Workplace. 2016;28:74-114.
13. Gray J. Infrastructural Experiments and the Politics of Open Access. En: Eve MP, Gray J (Eds.). Reassembling Scholarly Communications: Histories, Infrastructures and Global Politics of Open Access. The MIT Press;251-63. DOI: <https://doi.org/10.7551/mitpress/11885.001.0001>

14. Hagner M. Open Access, data capitalism and academic publishing. Swiss Med Week. 2018;148(78):1-8. DOI: <https://doi.org/10.4414/smw.2018.14600>
15. Haider J. Of the rich and the poor and other curious minds: On open access and “development”. Aslib Proceed. 2007;59(4/5):449-61. DOI: <https://doi.org/10.1108/00012530710817636>
16. Haider J. Openness as Tool for Acceleration and Measurement: Reflections on Problem Representations Underpinning Open Access and Open Science. En: Schöpfel J, Ulrich H. (Eds.). Open divide: Critical studies on open Access. Litwin Book; 2018. p. 17-28.
17. Hathcock AM. Preserving the Past for the Future: Whose Past? Everyone’s Future. En: Eve MP, Gray J (Eds.). Reassembling Scholarly Communications: Histories, Infrastructures, and Global Politics of Open Access. The MIT Press; 2020;205-14. DOI: <https://doi.org/10.7551/mitpress/11885.001.0001>
18. Holmwood J. Open Access, “Publicity,” and Democratic Knowledge. En: Eve MP, Gray J (Eds.). Reassembling Scholarly Communications: Histories, Infrastructures, and Global Politics of Open Access. The MIT Press. p. 181-90. DOI: <https://doi.org/10.7551/mitpress/11885.001.0001>
19. Herb U. Sociological implications of scientific publishing: Open access, science, society, democracy and the digital divide. First Mond. 2010;15(2). DOI: <https://doi.org/10.5210/fm.v15i2.2599>
20. Herb U. Open Access and Symbolic Gift Giving. En: Schöpfel J, Ulrich H. (Eds.). Open Divide? Critical Studies on Open Access. Litwin Books; 2018. p. 69-81. <https://www.zenodo.org/record/1206377>
21. Holley RP. Open Access: Current Overview and Future Prospects. Libr Trends. 2018;67(2):214-40. DOI: <https://doi.org/10.1353/lib.2018.0034>
22. Liyanage SS, MacIntyre CR. Do financial factors such as author page charges and industry funding impact on the nature of published

research in infectious diseases? Health Inform Libr J. 2006;23:214-22.

23. Mboa Nkoudou TH. Epistemic Alienation in African Scholarly Communications: Open Access as a Pharmakon. En: Eve MP, Gray J (Eds.). Reassembling Scholarly Communications: Histories, Infrastructures, and Global Politics of Open Access. The MIT Press; 2020. DOI: <https://doi.org/10.7551/mitpress/11885.001.0001>

24. Mirowski P. The future(s) of open science. Soc Stud Sci. 2018;48(2):171-203. DOI: <https://doi.org/10.1177/0306312718772086>

25. Moore SA. Open/Access: Negotiations around a concept. En: Schöpfel J, Herb H (Eds.). Open Divide? Critical Studies on Open Access. Litwin Books; 2018.

26. Piron F. Postcolonial Open Access. En: Schöpfel J, Herb H. (Eds.). Open Divide: Critical Studies on Open Access. Litwin Book; 2018. DOI: <https://scidecode.com/en/open-divide-critical-studies-on-open-access/>

27. Rév I. Accessing the Past, or Should Archives Provide Open Access? En: Eve MP Gray J (Eds.). Reassembling Scholarly Communications: Histories, Infrastructures, and Global Politics of Open Access. The MIT Press. 2020:229-47. DOI: <https://doi.org/10.7551/mitpress/11885.001.0001>

28. Roh C, Inefuku HW, Drabinski E. Scholarly Communications and Social Justice. En Eve MP Gray J (Eds.). Reassembling Scholarly Communications: Histories, Infrastructures, and Global Politics of Open Access. The MIT Press; 2020:41-52. DOI: <https://doi.org/10.7551/mitpress/11885.001.0001>

29. Shaw DM, Elger BS. Unethical Aspects of Open Access. Accountab Res. 2018;25(7-8): 409-16. DOI: <https://doi.org/10.1080/08989621.2018.1537789>

30. Schöpfel J. Open access-The rise and fall of a community-driven model of scientific communication. *Learn Publ.* 2015;28(4):321-5.

DOI: <https://doi.org/10.1087/20150413>

31. Šimukovič E. Open Access, a New Kind of Emerging Knowledge Regime? En: Schöpfel J, Herb H. (Eds.). *Open Divide: Critical Studies on Open Access.* Litwin Book; 2018. Disponible en:

<https://scidecode.com/en/open-divide-critical-studies-on-open-access/>

32. Tkacz N. From open source to open government: A critique of open politics. *Ephem: Theory Polit Organiz.* 2012;12(4). Disponible en:

<http://www.ephemerajournal.org/contribution/open-source-open-government-critique-open-politics-0>

33. Valsiner J. Open Access and its Social Context: New Colonialism in the Making? *Forum: Qualitative Social Research.* 2006;7(2):11.

### Conflicto de intereses

Se declara que no existe conflicto de intereses. La autora, *Nancy Sánchez Tarragó*, es una de las editoras de la Revista Cubana de Información en Ciencias de la Salud, pero no participó en la selección y designación de revisores, ni en ninguna otra etapa de revisión y edición de este artículo.