Herramienta de software para el análisis bibliométrico y de redes de producción científica

Software Tool for Bibliometric and Network Analyses of Scientific Production

Ferramenta de software para análise bibliométrica e redes de produção científica

Marilin Carolina Delfín Padrón
Centro de Investigación y Desarrollo en Tecnologías del Conocimiento (CIDTEC), Facultad de Humanidades y Educación, Universidad del Zulia. Maracaibo, Venezuela.
mardelfin@gmail.com

Gerardo Pirela
Laboratorio de Lenguajes y Modelos Computacionales, Facultad Experimental de Ciencias, Universidad del Zulia. Maracaibo, Venezuela.
gpirela@gmail.com

Resumen

El objetivo de este trabajo fue desarrollar una herramienta para la extracción automática y el análisis bibliométrico de redes de producción científica, la cual se realizó bajo la metodología incremental, generando dos grandes incrementos: la implementación de un algoritmo para la extracción automática de las redes de colaboración y referencia a partir de artículos científicos y el análisis bibliométrico de las redes de referencia y coautoría. Como resultado se obtuvo una herramienta para la extracción automática de redes de colaboración y referencia a través de técnicas de minería de texto aplicadas a artículos científicos digitalizados. Se concluyó que con este resultado se puede realizar estudios de tipo bibliométricos a la producción científica de LUZ. Se recomendó a las autoridades de Revicyh tomar en consideración permitir el acceso al repositorio de revistas, así como también la ampliación de esta herramienta a través de la inclusión de nuevas revistas.

Palabras clave: producción científica, bibliometría, minería de texto, redes de citación, redes de colaboración.

Abstract

The purpose of this research is to develop a software tool for automatic extraction and analysis bibliometric analysis of scientific production networks, which was held under the incremental methodology, generating two large increases: the implementation of an algorithm for automatic extraction of Reference and collaboration networks from scientific papers and the bibliometric analysis about reference networks and co-authorship. The result was a tool for automatic extraction of reference and collaboration networks through text mining techniques applied to digitized scientific papers. It was concluded that this result can be performed bibliometric studies on the LUZ scientific production. It recommended that the Revicyh authorities consider allow access to the repository of journals, as well as the expansion of this tool through the inclusion of new journals.

Keywords: text mining, network, collaboration, citation, bibliometrics.

Recibido: 2 de septiembre de 2017 Aprobado: 30 de noviembre de 2017

Cómo citar este artículo: Delfín Padrón M.C., Pirela G. (2017). Herramienta de software para el análisis bibliométrico y de redes de producción científica. *Códices, 13*(1), 109-125.

Resumo

O objetivo deste trabalho foi desenvolver uma ferramenta para a extração automática e análise bibliométrica de redes de produção científica, que foi realizada sob a metodologia incremental, gerando dois grandes incrementos: a implementação de um algoritmo para a extração automática das redes de colaboração e referência baseada em artigos científicos e análise bibliométrica das redes de referência e coautoria. Como resultado, foi obtida uma ferramenta para a extração automática de redes de colaboração e referência através de técnicas de mineração de texto aplicadas a artigos científicos digitalizados. Concluiu-se que com este resultado é possível realizar estudos bibliométricos para a produção científica de LUZ. Foi recomendado às autoridades do Revicyh que considerassem a permissão de acesso ao repositório de periódicos, bem como a extensão dessa ferramenta através da inclusão de novos periódicos.

Palavras chave: produção científica, bibliometria, mineração de textos, redes de citação, redes de colaboração.

Introducción

En los últimos años, el incremento de la producción científica ha demostrado un alto grado de multidisciplinariedad e interdisciplinaridad. El avance tecnológico ha permitido que las disciplinas se desarrollen vertiginosamente, por lo que hoy en día es más complejo representar a la ciencia y sobre todo establecer las relaciones que puedan tener una disciplina con otras.

Los instrumentos utilizados para medir los aspectos de la actividad científica son los indicadores bibliométricos: medidas que proporcionan información sobre los resultados de su producción (Castillo y Carretón, 2010) y que cuantifican el número de documentos publicados por país, institución, grupo de investigación o individuo; así como las citas recibidas por dichos documentos, entre muchos otros.

Actualmente, las redes sociales posibilitan el estudio de la ciencia a partir de las relaciones de sus componentes. Es decir, gracias a la teoría de grafos se hace posible la visualización de las relaciones entre disciplinas, autores, institutos de investigación y otros datos bibliográficos, obteniendo de esta manera alguna interpretación de lo que está sucediendo en el ámbito de la evolución de la investigación, la ciencia y las publicaciones científicas, transformando datos abstractos y fenómenos complejos de la realidad en mensajes visibles.

Pero, caracterizar la red según los tipos de elementos que la conformaría y representar la magnitud de la fuerza del vínculo que los une facilitaría

su análisis pero al mismo tiempo sumaría un grado de dificultad a lo antes mencionado.

En este sentido, es importante señalar la dificultad de procesamiento de este tipo de redes debido a la magnitud que pueden llegar a alcanzar y a la cantidad de información que se debe extraer y procesar para la creación de las mismas; de manera manual resultaría un proceso largo y engorroso, más aún realizar un análisis u obtener algún tipo de información significativa a partir de ella. Estos grafos son un gran aporte para la bibliometría ya que se puede determinar la emergencia de nuevos temas de investigación en el mundo así como también conocer la colaboración nacional o internacional de las instituciones, entre otros.

Por otra parte, la interpretación cuantitativa y cualitativa de los datos es uno de los factores clave que revelan el comportamiento de la comunidad científica en la investigación y publicación de artículos. Sin embargo, se debe realizar análisis multidimensional, en la que se evalúan simultáneamente varias variables como los indicadores métricos, la visualización mediante software empleado para el procesamiento de los datos y la interpretación de las estructuras de las redes sociales.

En términos de recuperación de información, el uso de interfaces gráficas para la representación de bases de datos bibliográficas se ha ido acentuando. Uno de los principales objetivos es lograr un modelo de representación que supere al tradicional bibliográfico referencial que, aunque útil, posee serias limitaciones al momento de mostrar al potencial usuario la verdadera estructura y dimensión del campo de conocimiento al cual se está enfrentando.

Un caso particular lo representa el Centro de Investigaciones y Desarrollo en Tecnologías del Conocimiento (CIDTEC) de La Universidad del Zulia (LUZ), ubicado en Maracaibo, Venezuela, en el que se hizo necesaria una herramienta de software que permita realizar un análisis cuantitativo de la producción científica. Esta herramienta resultó indispensable para la realización de sus propios estudios en cuanto al estado o crecimiento de la ciencia como resultado de investigaciones y publicaciones importantes, incluso poder medir el alcance que tiene las publicaciones realizadas dentro de dicha universidad así como una representación significativa de las relaciones anteriormente mencionadas.

El propósito de esta herramienta radica en facilitar la extracción, manipulación y representación gráfica de los elementos bibliográficos a partir de algunos artículos científico, específicamente publicaciones de la Universidad del Zulia, no todas las cuales se encuentran indexadas en las herramientas clásicas de acceso por la web para el tipo de análisis requerido.

Previo al desarrollo de la herramienta solicitada se realizó una revisión documental de herramientas existentes relacionadas a la requerida. Dos de estas herramientas sobresalieron en cuanto al análisis bibliográfico sobre la representación y análisis de redes de producción científica:

ARS Chile (2008), describe Ucinet como un paquete de software para el análisis de datos de redes sociales en la cual está incluida la herramienta de visualización red NetDraw. Ucinet puede trabajar con dos millones de nodos, sin embargo, en la práctica resulta bastante lento el proceso con 5.000 nodos, dependiendo del tipo de análisis que se desee realizar y al espesor de la red (cantidad de lazos o conectividad de la red).

Si bien Ucinet trabaja con todo tipo de redes, por esta misma generalidad, su algoritmo de minería de texto no escala adecuadamente con grandes bancos de datos cuando se quiere trabajar con uno u otro tipo de red específico.

Por su parte, HistCite, de Thomson Reuters (2011), es un software para el análisis y visualización de bibliografía de libre y gratuita distribución. HistCite ayuda a los investigadores con la visualización de los resultados de búsquedas *bibliográficas* en la *Web of Science,* permitiendo analizar y organizar los resultados de una búsqueda para obtener diferentes puntos de vista de la estructura del tema, la historia y las relaciones, proporcionando puntos de vista y de información no disponible de otra manera, como por ejemplo: autores y revistas más citados, productividad de las publicaciones y las tasas de citación dentro de una colección de documentos, relaciones de co-autoría, línea de tiempo de las publicaciones y visualización de histógrafos de línea de tiempo por campo de investigación.

HistCite es notoriamente un software bastante completo en cuanto al análisis bibliográfico y en cuanto al banco de datos que tiene a su alcance. Sin embargo, la producción científica de LUZ no se encuentra en ella ni permite agregar colecciones específicas a su banco de datos.

Redes y características topológicas

En matemáticas y ciencias de la computación, se define una red como un objeto llamado grafo y conformado por un conjunto de entidades (vértices) y las relaciones entre ellas (arcos o aristas), de la siguiente manera: G=(V,A) donde V es el conjunto de vértices (autores, artículos, contros de investigación, revistas, etc.) y A es el conjunto de arcos o aristas (relación de colaboración, de coautoría, de referencia, etc.).

En estos grafos, los vértices pueden conformarse y distribuirse en configuraciones particulares que permiten su estudio y caracterización. Por ejemplo: la distancia a la que están unos de otros, los apiñamientos en los que pueden agregarse ciertos subgrupos, la habilidad de formar cliques, el grado de interconexión local y global, entre otros factores pueden indicar qué tan robusta o vulnerable es la red, qué tan eficientemente puede fluir la información y otros tipos de señales a lo largo y ancho de la misma, así como otros indicadores tanto matemáticos como bibliométricos que ayudan a caracterizar la red, identificar potenciales debilidades y fortalezas e incluso apuntar a optimizar los recursos de la misma. Estas características se denominan topológicas por cuanto están íntimamente ligadas a las conformaciones geométricas y morfológicas de las redes en cuestión.

Distribución de grado

El grado de un nodo es una de las características topológicas más estudiadas. Martínez (2011, p. 26) define: «el grado de un nodo es el número de conexiones asociadas a un nodo. La distribución de grado se define como la distribución de probabilidad de un grado en un grafo». La distribución de grado es, entonces, la proporción de vértices con un valor de grado específico en relación a la cantidad total de vértices del grafo:

$$P(grado=g)\frac{\#\{vértices\,con\,grado=g\}}{\#\{vértices\,del\,grafo\}}$$

Ciertos grafos con características deseadas de robustez y flujo eficiente de información a lo largo de la red tienden a presentar ciertas distribuciones de grado específicas, las más comunes de las cuales se presentan a continuación:

- **Topología exponencial.** Martínez (2011) indica que esta distribución de grado Se produce en las redes evolucionistas en el tiempo, donde se posee la misma probabilidad de ser enlazado por el resto. A esto se le llama enlace igualitario. Es decir, una red que presente esta distribución de grados a lo largo del tiempo indicaría que sus entidades o vértices se están relacionando con una proporción o probabilidad casi constante: aproximadamente la misma cantidad o proporción de autores, coautores, referencias, correferencias, etc. están agregándose a la red dentro de lapsos de tiempo más o menos constantes. La ecuación de la distribución de grado exponencial es:

$$P(k) = Ce^{\alpha k}$$

  en la que los valores C y α se denominan parámetros de distribución, son particulares a cada red de este tipo e indican la proporción y velocidad de agregación de entidades y relaciones a ella.
- **Topología libre de escala.** Según Mitchell (2009), en este tipo de redes complejas (redes de libre escalamiento), existen un número reducido de nodos que tienen un grado alto (denominados ***hubs*** o concentradores) y muchos nodos que están pobremente conectados con el resto de los nodos del grafo (denominados nodos satelitales), por lo que la distribución de grado sigue una ley de potencias:

$$P(k) = Ck^{-\gamma} \mid \gamma = 0$$

  donde los parámetros C y γ son particular para cada red e indican el efecto de nodos *hubs* sobre los satelitales.

De algunos estudios se desprende que ésta es la distribución que mejor describe muchos fenómenos naturales y sociales que muestran ser robustos a errores aleatorios. Esta distribución aparece en redes complejas de distintos dominios como: redes neuronales en el cerebro de muchos animales, redes de distribución eléctrica en zonas urbanas y redes sociales (entre otras). En éstas, el valor típico del parámetro γ oscila entre 2 y 3. Por ser las redes estudiadas por la bibliometría casos particulares de redes sociales, se espera que la distribución de grado en estas redes sea de este tipo.

Distancia entre nodos

De acuerdo a Bouttier y col. (2003), se denomina distancia entre dos nodos de un grafo al número de nodos mínimo que debe recorrerse para unirlos. La distancia entre dos nodos de un grafo es la longitud del camino más corto (un camino es cualquier ruta entre dos nodos donde ningún nodo es visitado más de una vez). Si no hubiera conexión alguna entre dos nodos se dice que la distancia es infinita. Las distancias de todos los nodos de un grafo se computan en lo que se denomina matriz de distancias. El concepto se emplea en las mediadas de centralidad de redes.

Excentricidad

Para Cook (2014), la excentricidad de un nodo es la distancia más larga a partir de ese nodo a cualquier otro nodo en la red. Los nodos con menor excentricidad son más centrales, mientras que los de mayor excentricidad son más periféricos.

Coeficiente de apiñamiento

Según lo indicado por Martínez (2011), se define el coeficiente de apiñamiento, *clustering* o coeficiente de agrupamiento de un nodo i como la cantidad de enlaces existentes entre sus vecinos dividido entre la cantidad total de enlaces que pudiera haber entre ellos. Dado un grafo $G = (V,E)$, el coeficiente de apiñamiento C_i se define como:

$$C_i = \frac{2\tau_i}{g_i(g_i - 1)}$$

donde τ_i es la cantidad de triángulos en los que está involucrado el nodo *i* (la cantidad de conexiones entre sus vecinos) y g_i es el grado del nodo *i*. Valores altos de este indicador para un nodo particular indica cuán bien conectado está dicho nodo en la red: cuan robusta es la red alrededor de dicho nodo respecto a potenciales pérdidas de conexiones individuales (si los vecinos están bien conectados y se pierde uno o dos conexiones, la información seguirá fluyendo por rutas alternativas); en contraposición, si un nodo tiene bajo coeficiente de apiñamiento, su pérdida podría representar falla en el flujo de

información a lo largo de la red desde este nodo a algunos de sus vecinos (indicando vulnerabilidad a ataques sobre la red alrededor de este nodo).

La evidencia sugiere que en la mayoría de redes del mundo real, y en particular las redes sociales, los nodos tienden a crear grupos muy unidos que se caracterizan por una densidad relativamente alta de enlaces.

Análisis bibliométrico

El análisis bibliométrico es un método documental que ha alcanzado un importante desarrollo durante las últimas décadas. «Sus objetivos fundamentales son, por una parte, el estudio del tamaño, crecimiento y distribución de los documentos científicos y, por otra, la indagación de la estructura y dinámica de los grupos que producen y consumen dichos documentos y la información que contienen» (González y col., 1997, p. 236).

Indicadores bibliométricos

El análisis bibliométrico se realiza a través de indicadores bibliométricos, definidos como los instrumentos para medir la producción científica permitiendo conocer el impacto causado por un trabajo científico cualquiera a partir de la literatura científica y tecnológica publicada. Los indicadores bibliométricos permiten manejar, clasificar y analizar grandes volúmenes de publicaciones científicas (Ruiz, 2005).

Visto de esta manera, los indicadores son «los parámetros que se utilizan en el proceso evolutivo de cualquier actividad» (Sancho, 1990, p.843). Normalmente se emplea un conjunto de ellos, cada uno de los cuales pone de manifiesto una característica del objeto de estudio. Esto se hace evidente en el caso de la ciencia y de las publicaciones, que al ser multidimensionales, es necesario calcular distintos tipos de indicadores.

Con los indicadores bibliométricos se podrán determinar, entre otros, los siguientes aspectos:

- El crecimiento de cualquier campo de la ciencia, según la variación cronológica del número de trabajos publicados en él.
- El envejecimiento de los campos científicos, según la «vida media» de las referencias de sus publicaciones.

- La evolución cronológica de la producción científica, según el año de publicación de los documentos.
- La productividad de los autores o instituciones, medida por el número de sus trabajos.
- La colaboración entre los científicos e instituciones, medida por el número de autores por trabajo o centros de investigación que colaboran.
- La dispersión de las publicaciones científicas entre las diversas fuentes.

Según González y col. (1997), en la última década, la cantidad de publicaciones se ha visto en aumento acompañado de un progresivo incremento de la calidad y del impacto de las publicaciones. Por otra parte, desde una perspectiva histórica y sociológica, la participación de varios autores en la elaboración de un trabajo es consecuencia de la profesionalización de la comunidad científica. A principios de siglo el 80% de los trabajos científicos tenían una sola firma es decir, contaban con un solo autor, mientras que en la actualidad aproximadamente el 80 % tienen varias firmas.

La proporción de artículos firmados por varios autores aumenta cuando se trata de trabajos que reciben ayuda económica, lo que apoya la relación entre colaboración y soporte financiero. La cuestión del orden de firma de los autores es también compleja. Mientras que lo más usual es que firme en primer lugar el investigador principal, el orden de los siguientes no refleja necesariamente el grado de colaboración.

Por otro lado, también se encuentra el análisis de referencias comunes donde, si dos publicaciones poseen una o más referencias comunes se puede decir que están bibliográficamente relacionados y, por tanto, pertenecen al mismo campo de conocimiento. Cuantas más referencias comunes aparecen en los trabajos, más cercana será su temática.

Materiales y métodos

Para el desarrollo de la herramienta planteada se utilizó la metodología incremental de desarrollo de software propuesta por Harlan Mills en el año 1980 y descrita por Bermúdez y col. (2011) como un enfoque evolutivo que combina el modelo lineal con la interacción del modelo por prototipado, ejecutándose de manera iterativa (escalonada) con el objetivo de que en cada iteración

se obtenga un «incremento»: un módulo o parte funcional del sistema final de manera que el usuario o cliente pueda dar revisiones y correcciones sobre cada incremento minimizando de esta forma los errores y adaptándola mejor a sus necesidades reales.

Debido a la naturaleza de la herramienta desarrollada, se definieron dos grandes incrementos, cada uno de los cuales constó de las etapas clásicas del ciclo de vida de un software: análisis, diseño, implementación y prueba. Estos incrementos fueron:

Implementación de un algoritmo para la extracción automática de las redes de colaboración y referencias a partir de artículos científicos digitalizados

- Análisis: se identificaron los datos bibliográficos que se debían extraer y se seleccionaron las revistas a utilizar para los casos de estudio o prueba de principio, para de éstas identificar los formatos de los datos bibliográficos a ser extraídos (título de artículos, nombres de autores, referencias bibliográficas, etc.). Además, se realizó la selección de la técnica de minería de texto para la extracción automática de los datos bibliográficos necesarios.
- Diseño: se realizó el diseño para las estructuras de datos que representan la red de colaboración y de referencias.
- Implementación: se implementó el algoritmo de minería de texto con el cual se extrajo de manera automática las redes de colaboración y de referencias de los artículos científicos digitalizados.
- Prueba: se realizaron las pruebas para comprobar el buen funcionamiento del algoritmo anteriormente mencionado y de esta manera comprobar que realizaba de manera satisfactoria la extracción de la red de colaboración y de referencia.

Análisis bibliométrico de las redes de colaboración y de referencias

- Análisis: se seleccionaron los indicadores para el análisis bibliométrico.
- Diseño: se realizó el bosquejo o diseño de lo que sería el algoritmo de análisis bibliométrico aplicado a las redes.
- Implementación: Se realizó la implementación del algoritmo para el análisis bibliométrico en la red extraída.

- Prueba: Se realizaron las pruebas pertinentes para comprobar el correcto funcionamiento del algoritmo para el análisis bibliométrico de las redes de colaboración y de referencias.

Cabe destacar que las pruebas de principio se llevaron a cabo basándose en el análisis de cinco revistas de la producción científica de LUZ, específicamente: *Enl@ce, Lingua Americana, Quórum Académico, Revista de Literatura Hispanoamericana y Revista de Filosofía*. La idea original era que el programa se conectara *directamente* al repositorio de la producción científica de LUZ Revicyh y de esta manera tener mayor alcance de los insumos de datos permitiendo una actualización mucho más rápida. A pesar de que la permisología fue negada por el organismo correspondiente, esto no fue un altercado significativo para continuar con el proceso, ya que se implementó un banco de datos local para las publicaciones digitalizadas a partir de las cuales se extraerán los datos pertinentes luego de pasar por el proceso de reconocimiento óptico de caracteres. La revista *Lingua Americana* fue seleccionada arbitrariamente como caso de estudio para realizar las pruebas pertinentes y verificar el buen funcionamiento de la herramienta.

Resultados

Arquitectura de BiblioRed

La herramienta de software construida se denominó BiblioRed y su arquitectura funcional general se muestra en la imagen 1:

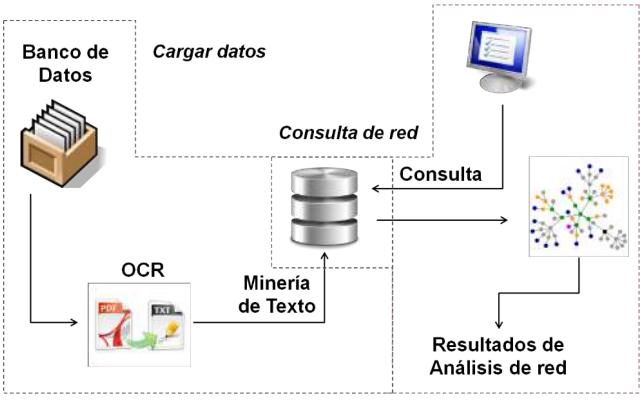

Imagen 1. Arquitectura de la herramienta.

Se tiene un banco de datos para las revistas ***digitalizadas*** en formato. PDF el cual fue descargado de manera manual a partir del portal de Revicyh. Luego, se realiza el reconocimiento óptico de caracteres para convertir los artículos de formato.PDF a formato.TXT con la ayuda de un programa externo llamado *Some PDF to Txt Converter*, (software gratuito, rápido y que permite la carga de múltiples archivos). Una vez obtenidos los archivos.TXT es posible ejecutar el algoritmo de minería de texto sobre éstos.

El algoritmo de minería de texto funciona de la siguiente manera:

El archivo está dividido en cinco partes: cabecera, título, autores, contenido y referencias bibliográficas, tal que de la cabecera se obtienen datos como el nombre de la revista, año, volumen y número al cual pertenece ese artículo.

Se utiliza la técnica de reconocimiento de patrones textuales basada en expresiones regulares, para lo cual se definieron las siguientes:

- Nombre de la revista: [a-zA-ZáéíóúÁÉÍÓÚñÑ']+
- Volumen: [I|V|X|L|C|D|M]+
- Año:[0-9]+
- Número:[0-9]+

Para las secciones sucesivas, las expresiones regulares se definieron de la siguiente manera:

- Título: [a-zA-ZáéíóúÁÉÍÓÚñÑ0-9α]+ donde, α={-,.,:,",',!,¡,?.¿}
- Autores: Para un autor [a-zA-ZáéíóúÁÉÍÓÚñÑ']+[[a-zA-ZáéíóúÁÉÍÓÚñÑ'].] ? [a-zA-ZáéíóúÁÉÍÓÚñÑ']?
- Para varios autores: Ω= autor [a-zA-ZáéíóúÁÉÍÓÚñÑ']+ [[a-zA-ZáéíóúÁÉÍÓÚñÑ'].] ? [a-zA-ZáéíóúÁÉÍÓÚñÑ']? entonces, Ω (,|y|e Ω)+
- Referencias bibliográficas:
  - Autores [a-zA-ZáéíóúÁÉÍÓÚñÑ']+, [a-zA-ZáéíóúÁÉÍÓÚñÑ']+[; [a-zA-ZáéíóúÁÉÍÓÚñÑ']+, [a-zA-ZáéíóúÁÉÍÓÚñÑ']+]?
  - Año: [0-9]+
  - Título: [a-zA-ZáéíóúÁÉÍÓÚñÑ0-9α]+ donde, α={-,.,:,",',!,¡,?.¿}

Cada uno de los datos extraídos es almacenado en una base de datos diseñada bajo las normalizaciones correspondientes y de acuerdo a los requerimientos de la herramienta. Esta base de datos cuenta con siete tablas como se puede observar en la imagen 2 con más detalle.

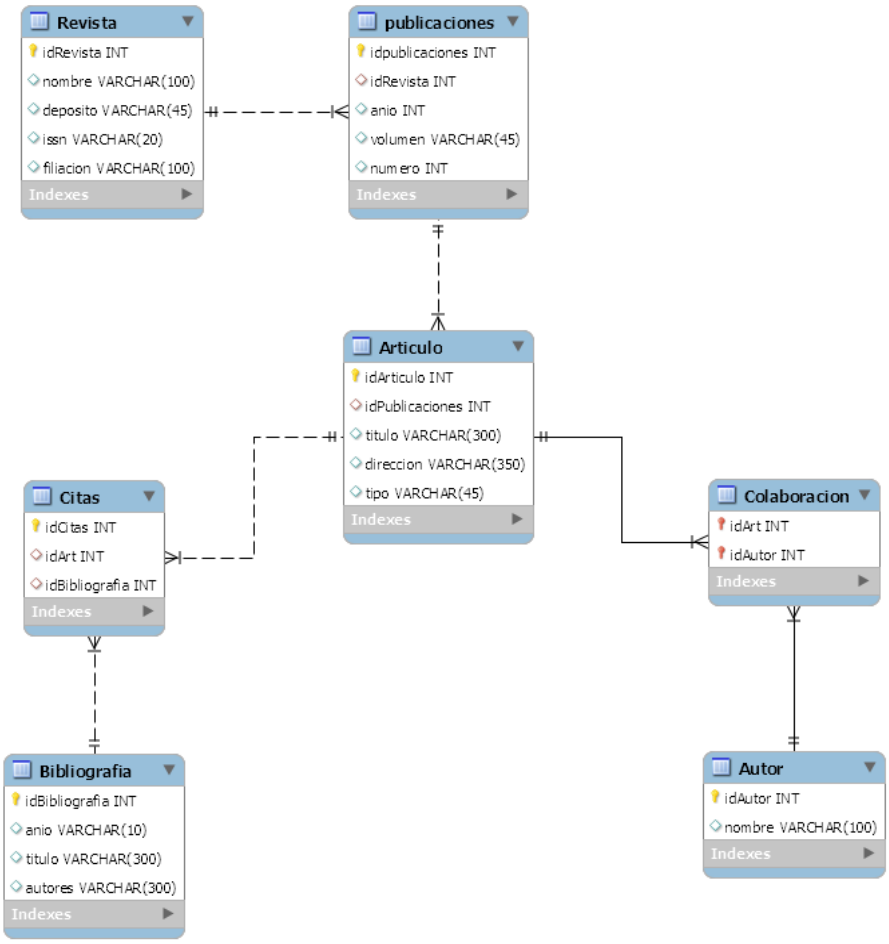

Imagen 2. Diagrama relacional de la base de datos.

Para la consulta o generación de una red se implementó una interfaz con parámetros como selección de revistas, el período que abarcará y el tipo de red que se desea consultar: de colaboración o de referencias. Al procesar la consulta en la base de datos se extraerán los datos necesarios para dibujar la red mediante la implementación del *toolkit* de Gephy (un paquete java para la visualización de redes). Seguidamente, los datos de la red consultada son cargados en las estructuras computacionales implementadas para tal fin, llamadas: `Matriz_ady` (donde se realizan los cálculos para las características topológicas de la red: distancias, excentricidades, coeficientes de apiñamiento

y la excentricidad) y *Bibliometrico* (que se encarga de calcular los indicadores bibliométricos correspondientes).

La estructura computacional que implementa el algoritmo de minería de texto se llama *Archivo*, en la que, para mayor simplicidad, se implementó un método por cada tipo de documento y revista que se esté procesando (por ejemplo: artículos, reseña, enciclopedia, entre otros). Estos métodos procesan el archivo para la extracción de datos según el formato del tipo de documento y según las expresiones regulares descritas anteriormente. Así mismo, cada método llama a otro llamado *bibliografía()* para procesar las referencias bibliográficas de acuerdo al formato en el que se encuentren (según las expresiones regulares que se hayan definido para tal patrón).

Para la visualización de la red se emplea la estructura *grafo*, en la que, a través del método *script()*, se realiza la consulta a la base de datos de acuerdo a los parámetros elegidos y se construye la red para su posterior visualización.

Pruebas funcionales de BiblioRed

Con el fin de corroborar la funcionalidad de la herramienta se realizaron pruebas de tipo caja negra, tomando *Lingua Americana* como revista piloto. En la imagen 3 se ilustra la prueba de carga satisfactoria de artículos, indicando que fue ejecutado el algoritmo de minería de texto y los datos extraídos fueron almacenados en la base de datos.

De la misma manera, fueron realizadas pruebas para la consulta de red de manera satisfactoria, tal como se ilustra en la imagen 4, mediante la cual se obtuvo la representación gráfica de la misma en forma de grafo y los indicadores bibliométricos correspondientes.

Conclusiones

A pesar de que existen herramientas con propósitos semejantes a los planteados para BiblioRed, como lo son la visualización de redes y el análisis bibliométrico, no existen herramientas de extracción automáticas de redes a partir de artículos digitalizados propios, locales o a través de conexiones a bases de datos

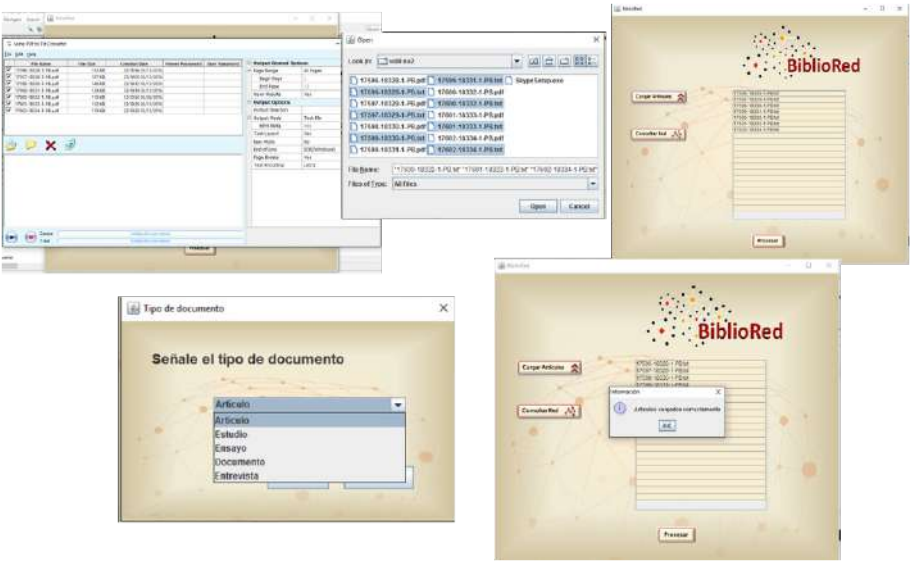

Imagen 3. Prueba de carga de artículo.

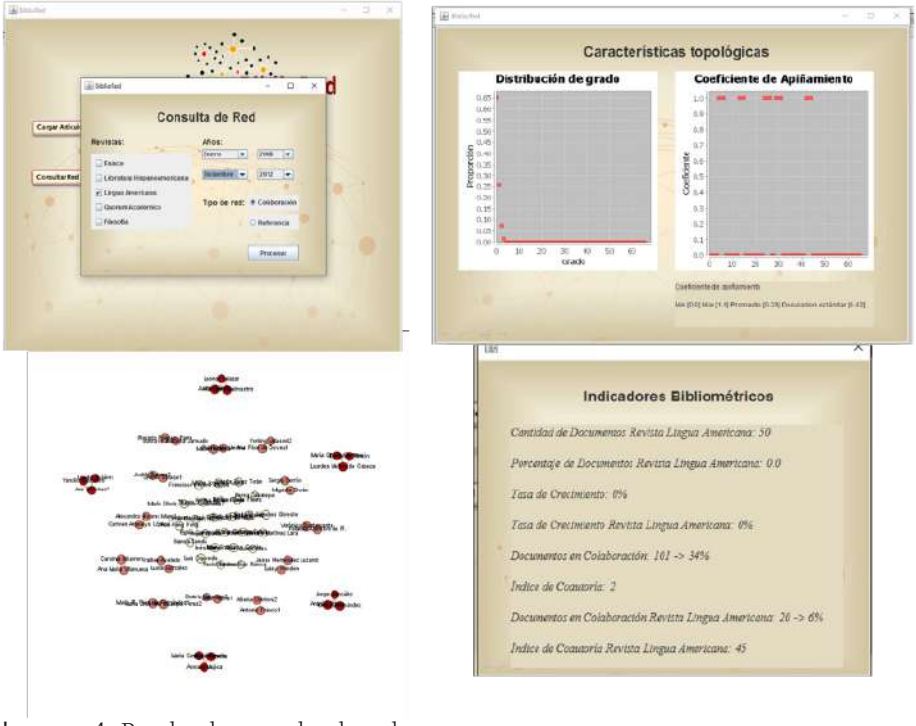

Imagen 4. Prueba de consulta de red.

de particular interés para estudios específicos, como fue el caso del repositorio de Revicyh que contiene la producción científica de La Universidad del Zulia, el estudio bibliométrico de la cual no se había podido realizar hasta ahora con el enfoque de redes e incorporando indicadores topológicos de la misma.

La implementación de una algoritmo de minería de texto basado en patrones al estilo de expresiones regulares aplicadas a archivos en formato de texto plano y el uso de tecnología libre para la transformación de documentos digitalizados en formato PDF a documentos de texto plano permitió la independización funcional del algoritmo para la extracción automática de redes de la conexión a base de datos. Esto fue particularmente ventajoso, ya que no se cuenta aún con permisología suficiente para acceder al amplísimo repositorio de producción científica de LUZ; sin embargo, el sistema es suficientemente flexible y versátil como para funcionar correctamente dada la ubicación física de un banco de documentos PDF conteniente de los artículos que se desea analizar y la correcta configuración de las expresiones regulares correspondientes a las secciones de estos artículos.

Los autores esperan que los resultados obtenidos hasta ahora sirvan para estimular estudios sucesivos en las líneas de investigación ***transdisciplinarias*** de Teoría de Grafos (por parte de las ciencias de la computación) y Bibliometría (por parte de las ciencias de la información). Así mismo, se espera contar con insumos suficientemente masivos que permitan la evaluación de la escalabilidad del software resultante y la comprobación de principios topológicos como la caracterización de redes de producción científica en la categoría de redes de libre escalamiento (con distribución de grados de libre escala y altos valores de coeficientes de apiñamiento que indiquen robustez, eficiente flujo de información a lo largo de la red y resistencia a ataques).

Referencias

ARS Chile (2008) *Primeros pasos con Ucinet 6*. (Página consultada el 20 de mayo de 2014) Dirección URL: http://www.arschile.cl/ucinet/index.html

BERMÚDEZ, Cristian; GARRIDO, Erika; LARA, Natalia (2011). *Metodología Incremental*. (Página consultada el 20 de mayo de 2014) Dirección URL: https://procesosoftware.wikispaces.com/Modelo+Incremental

BOUTTIER, Jérémie; DI FRANCESCO, P., GUITTER, E. (2003). «Geodesic distance in planar graphs» en *Nuclear Physics B*. Volumen 663, p. 535–567

CASTILLO, A.; CARRETÓN, M. (2010). Investigación en Comunicación. Estudio bibliométrico de las Revistas de Comunicación en España. en: *Comunicación y Sociedad,* vol. XXIII, n. 2, 2010, pp.289-327. [en línea] http://www.ecured.cu/index.php/Bibliometr%C3%ADa Consultado el 14/06/2013

COOK, Samantha (2014). *Medidas de centralidad.* (Página consultada el 10 de enero de 2016) Dirección URL: http://ars-uns.blogspot.com/2014/01/ars-101-medidas-de-centralidad-nuevas.html

GONZÁLEZ de Dios, J; MOYA, M; MATEOS HERNÁNDEZ, M. (1997). Indicadores bibliométricos: Características y limitaciones en el análisis de la actividad científica. Artículo especial. Vol. 47. No. 3. Madrid. España.

MITCHELL, Melanie (2009). *Complexity – A guided tour.* Oxford University Press. New York. USA.

MARTÍNEZ ARQUÉ, Néstor (2011). *Análisis, comparativa y visualización de redes sociales on-line representadas como grafos* (Página consultada el 10 de enero de 2016) Dirección URL: https://repositori.upf.edu/bitstream/handle/10230/12978/PFC_Nestor_Martinez.pdf?sequence=1

RUIZ DE OSMA, Elvira (2005). *Indicadores bibliométricos.* (Página consultada el 25 de enero de 2016) Dirección URL: http://www.ugr.es/~rruizb/cognosfera/sala_de_estudio/ciencimetria_redes_conocimiento/indicadores_bibliometricos.htm

SANCHO CAPARRINI, Fernando (2015). *Introducción a las redes complejas.* (Página consultada el 10 de enero de 2016) Dirección URL: http://www.cs.us.es/~fsancho/?e=80

SANCHO, Rosa (1990). *Indicadores bibliométricos utilizados en la evaluación de la ciencia y la tecnología.* Madrid. España.

THOMSON REUTERS (2011). *HistCite*. (Página consultada el 24 de mayo de 2013) Dirección URL: http://thomsonreuters.com/products_services/science/science_products/a-z/histcite/