



---

## NOTAS Y EXPERIENCIAS / NOTES AND EXPERIENCES

---

### **“Desktop Scientometrics”: una metodología para el uso de datos procedentes de WoS mediante el programa estadístico R**

Preiddy Efrain-García\*, Carlos García-Zorita\*\*

\*Centro de Ciencias Humanas y Sociales (CCHS), Consejo Superior de Investigaciones Científicas (CSIC). Madrid, España.

\*\*Laboratorio de Estudios Métricos de la Información (LEMI), Departamento de Biblioteconomía y Documentación.

Universidad Carlos III de Madrid.

Correo-e: [preiddy@gmail.com](mailto:preiddy@gmail.com)

Recibido: 11-09-2014; 2ª versión: 05-05-2015; Aceptado: 22-05-2015.

**Cómo citar este artículo/Citation:** Efrain-García, P. y García-Zorita, C. (2016). “Desktop Scientometrics”: una metodología para el uso de datos procedentes de WoS mediante el programa estadístico R. *Revista Española de Documentación Científica*, 39(1): e122. doi: <http://dx.doi.org/10.3989/redc.2016.1.1223>

**Resumen:** Se presenta una metodología de trabajo que permite obtener y tratar datos de la *Web of Science* con fines bibliométricos. Se usa como herramienta el software estadístico R para crear funciones que faciliten el trabajo bibliométrico; adicionalmente se genera una base de datos relacional que potencia la gestión de los datos y la generación de indicadores bibliométricos.

**Palabras clave:** Métodos bibliométricos; bases de datos relacionales; SQL; R; Web of Science; indicadores bibliométricos.

#### **“Desktop Scientometrics”: a methodology using the R statistical software with data from WoS**

**Abstract:** A methodology is presented for obtaining and processing data from the Web of Science (WoS) for bibliometric purposes. The R statistical software is used as a tool for creating functions that facilitate bibliometric work. In addition, a relational database was created that enhances both data management and the generation of bibliometric indicators.

**Keywords:** Bibliometric methods; relational databases; SQL; R; Web of Science; bibliometric indicators.

**Copyright:** © 2016 CSIC. Este es un artículo de acceso abierto distribuido bajo los términos de la licencia Creative Commons Attribution-Non Commercial (by-nc) Spain 3.0.

## 1. INTRODUCCIÓN

Desde que J.S. Katz y Diana Hicks (1997) propusieran el término *Desktop Scientometrics* se ha ido consolidando un "concepto" o "filosofía" de trabajo en relación con los estudios (biblio-) métricos que se ha caracterizado por el uso de herramientas informáticas personales (hardware y software) cada vez más potentes y por la creciente disponibilidad de bases de datos susceptibles de uso bibliométrico.

En estos años se ha llevado a cabo un proceso paralelo en relación con la aparición y desarrollo de aplicaciones de *software libre*, generalmente disponibles en formato multiplataforma, muy accesibles y suministradas en diversos niveles de implementación que posibilitan su uso por diferentes tipos de usuarios.

En este sentido, en el ámbito de los estudios métricos de información es habitual el uso de lenguajes de programación como *Perl* (García González, 2010) o *Python* (Grauwin y Jensen, 2011), la utilización de gestores de bases de datos como *MySQL* (Anuradha y Urs, 2007; Han y Sun, 2014), *Postgre* (Bowman y otros, 2014) o *SQLite* (Gagolewski, 2011), o los análisis llevados a cabo con aplicaciones como *Pajek* (Zhang, y otros, 2013), *CiteSpace* (Chen y otros, 2012) o *Gephi* (Leydesdorff y Rafols, 2012). Todas estas aplicaciones están disponibles para su uso libre y suponen una nueva forma de entender la investigación en la que este trabajo se inscribe.

Por otra parte, la profesión de bibliotecario y en particular la del bibliotecario universitario, se está adaptando en sus competencias a nuevos escenarios. En el caso de las bibliotecas universitarias se está notando un cierto interés en la forma en la que los bibliotecarios pueden integrarse de forma colaborativa en las instituciones que las acogen. Conceptos como el de *'embedded librarian'* (Dewey, 2004) abogan por una mayor integración y colaboración de los bibliotecarios en las tareas de enseñanza-aprendizaje y en las labores de investigación. Tener *'implicados'* profesionales de la información en estos ámbitos académicos puede reportar beneficios a investigadores (Carlson y Kneale, 2011) y docentes (Hernández-Pérez y otros, 2011) para crear nuevas oportunidades a los profesionales (Kesselman y Watstein, 2009).

Así mismo, el creciente interés de las Universidades en las comparaciones nacionales e internacionales *-rankings-* y en la institucionalización de la evaluación de sus profesores en su faceta de investigadores, como ocurre con las acreditaciones o los sexenios en España o el *Research Assessment Exercise* (RAE) en el Reino Unido, han

conducido a un creciente interés en las bibliotecas universitarias por la bibliometría y los servicios que ésta puede prestar en las actividades de evaluación de la investigación de estas instituciones (Corrall y otros, 2013).

Esta nueva apuesta por las competencias bibliométricas / informáticas de los bibliotecarios se ha resuelto incluso contratando expertos en esta área<sup>1</sup> que en el caso de la Universidad de Leicester se hizo incluyendo un bibliómetra en el servicio bibliotecario, sin apostar por la reconversión de sus bibliotecarios (Corrall y otros, 2013).

Sin embargo, en nuestra opinión, y este es el objetivo del presente trabajo, es factible dotar a los profesionales de las bibliotecas universitarias y de investigación de herramientas sencillas que les permitan adquirir y desarrollar competencias básicas que les den la posibilidad de abordar estudios métricos de información desde cualquier computador personal, mediante una metodología basada en el uso de software libre y sin pérdida de control sobre los datos a analizar.

## 2. BIBLIOMETRÍA CON R

El trabajo bibliométrico que se lleva a cabo hoy en día requiere el empleo de competencias multidisciplinarias. Además de las propias de la metodología bibliométrica y como ya se ha insinuado más arriba, son necesarias competencias estadísticas e informáticas. En nuestra propuesta desarrollamos una metodología de tratamiento y análisis de datos que trata de combinar algunos aspectos de estas dos últimas competencias.

R es un conjunto integrado de librerías y un lenguaje de programación que nos permite realizar cálculos fundamentalmente estadísticos (R Core Team, 2012).

R es un entorno en el que se encuentran implementadas muchas técnicas y métodos estadísticos, que van desde los modelos lineales a las series temporales, pasando por los test clásicos. El conjunto básico de instalación se encuentra compuesto por ocho librerías que son llamadas "estándar" (Arriaza Gómez y otros., 2008; Paradis, 2005). Se puede descargar desde su página web principal: <http://www.r-project.org>. Es software libre y se encuentran versiones para los sistemas operativos más empleados en la actualidad<sup>2</sup>.

La comunidad de colaboradores que utilizan el entorno R ha desarrollado centenares de aplicaciones (*packages*) que cubren casi cualquier necesidad de análisis y tratamiento estadístico de datos. En este sentido, y en el ámbito bibliométrico, Gagolewski (2011) ha desarrollado un paquete (CITAN)

para el análisis bibliométrico de citas a partir de datos bibliográficos obtenidos de *Sciverse-Scopus*.

Nuestra intención ahora no es desarrollar un paquete *ad-hoc* para el análisis bibliométrico, sino mostrar, a través de la presentación de algunas funciones sencillas y eficaces, que es relativamente fácil adquirir las competencias necesarias para llevar a cabo análisis más completos.

El entorno de funcionamiento del programa se basa en que posee diversas estructuras de datos denominadas "objetos", a los cuales se le suelen asignar nombres, contenidos y atributos. A continuación mencionamos brevemente algunos de estos objetos:

- *Vector*: Es la estructura más simple, cuenta con dos argumentos *mode* y *length*. Los vectores pueden ser de tres tipos, a saber: numérico, lógico o de carácter. Un vector solo posee un tipo de dato. (R Core Team, 2012; Paradis, 2005).
- *Factor*: "Un factor incluye no solo los valores correspondientes a una variable categórica, sino también los diferentes niveles posibles de esta variable" (Paradis, 2005).
- *Matriz*: Son vectores que cuentan con un atributo adicional (*dim*), que a su vez es un vector numérico de longitud 2 que define el número de las filas y columnas que deberá tener. Las matrices también son denominadas variables indexadas o *arrays* (R Core Team, 2012).
- *Listas*: es un tipo de objeto que está conformado por vectores de diferentes tipos.
- *Dataframes*: son el símil de las bases de datos o las hojas de cálculo. Pueden estar conformadas por diferentes tipos de datos como las listas y más de una columna como en una matriz.

A todos los objetos antes descritos se les pueden aplicar funciones: desde las sencillas de carácter aritmético hasta complejas funciones para el análisis multivariable. La única preocupación que hay que tener presente siempre es con qué tipo de objeto se está trabajando para así poder determinar qué tipo de función podemos aplicar.

Se entiende como función un conjunto de instrucciones escritas en R que reciben un objeto y lo procesan de acuerdo a esas instrucciones. En realidad, todo R trabaja con funciones, es decir, todas las instrucciones que ejecutamos en R son funciones. Para hacernos una idea de qué son las funciones lo mejor es hacer un símil con las macros que se pueden crear en las hojas de cálculo (Paradis, 2005).

Por ser R un lenguaje de programación, el usuario puede programar sus propias funciones. Para ello hay que tener en cuenta algunos aspectos:

- No es necesario *declarar* las variables dentro del entorno de la función, como se hace en otros lenguajes.
- Se escriben directamente en el entorno de trabajo o en un editor externo de texto ASCII.
- Si son escritas directamente en el entorno de trabajo son almacenadas en memoria.
- Si las funciones son escritas con editores externos solo hay que emplear la instrucción *source* (nombre-función) para cargarla en memoria.

En la metodología que aquí presentamos, todos los pasos necesarios para cargar los datos y obtener algunos indicadores han sido programados como una función. Se ha realizado de esta manera para que el usuario final se concentre en lo que consideramos realmente importante, el análisis de los datos y no en aspectos más técnicos. Sin embargo, todo el código de las funciones que empleamos se encuentran en texto plano y pueden ser modificadas y mejoradas con el único compromiso de comunicarlo a los autores.

Ilustramos con un ejemplo la creación de una función y seguidamente se comentan las tareas que realiza.

```
mifuncion=function(x){datos= read.table(x)
plot(datos$V1)}
```

En la primera línea se declara el nombre que tendrá la función (*mifuncion*) para seguidamente entrar en el entorno *function*. En la segunda línea se indica al programa que lea un archivo de texto cuyo nombre se ha pasado como valor (*x*) de la función, el comando *read.table* lee el fichero y su contenido lo almacena en una variable, en este caso del tipo *dataframe*, denominada *datos*. La última línea realiza la correspondiente gráfica (*plot*) de la primera variable (columna *V1*) del archivo indicado en el argumento.

No se pretende que el usuario sea un programador experimentado, sino que disponga de las nociones básicas del funcionamiento de R y que en la medida que el modelo de trabajo que proponemos le resulte útil y motivador, se anime a modificar y crear sus propias funciones.

### 3. MATERIAL Y MÉTODO

En el método bibliométrico habitualmente son dos las fases a las que hay que enfrentarse: la preparación de los datos y la obtención de indicadores.

Mostramos a continuación cómo afrontar estas dos fases utilizando un conjunto de funciones programadas en el entorno R. El lector podrá ejecutarlas simplemente y si lo desea revisarlas y adaptarlas a sus necesidades.

### 3.1. Preparación y Obtención de los Datos

En esta primera fase, han de seguirse los pasos necesarios para la obtención y el posterior tratamiento de los datos bibliográficos susceptibles de uso bibliométrico. Una vez seleccionada la fuente de información y obtenidos los datos objeto del estudio es aconsejable la carga, tratamiento e integración de los datos en un sistema de gestión de bases de datos, preferentemente de carácter relacional. Nuestra elección en este trabajo es la *BD SQLite* debido a su sencillez de manejo (un único fichero por cada base de datos). Gestores de *SQLite* hay muchos disponibles libremente y para todos los sistemas operativos, aunque por su simplicidad de instalación recomendamos el complemento disponible para *Firefox*<sup>3</sup>. Esta fase debe concluir finalmente con la depuración y homologación de los datos.

#### i. Selección

En el presente trabajo y para ilustrar la metodología propuesta se utilizarán los índices de citas de la *Web of Science (Thomson-Reuters)*; los resultados se almacenarán en archivos de texto plano con extensión “.txt”. Se ha de seleccionar el formato delimitado por tabuladores. Para evitar, en la medida de lo posible, errores en la importación y por la experiencia de trabajo con plataformas Windows y Mac, sugerimos que los datos sean exportados con el tabulado correspondiente a la plataforma que estemos usando y con la codificación de caracteres UTF-8.

#### ii. Carga de los datos

Para ilustrar las distintas fases metodológicas, y a modo de ejemplo, en este trabajo se ha trabajado con una sola descarga de 500 registros del *Social Sciences Citation Index (SSCI- WoS)*, pero el usuario puede trabajar con el número de ficheros que desee.

Una vez obtenido el o los archivos con el formato delimitado por tabuladores, iniciamos R y procedemos a cargar las funciones que hemos desarrollado, las cuales se encuentran disponibles en la siguiente dirección web: [https://github.com/06122010/r\\_isi](https://github.com/06122010/r_isi)<sup>4</sup>

Procedemos a cargar en el entorno de trabajo en R las funciones. A continuación mostramos la instrucción para cargar todas las funciones (no olvide la ruta de acceso al fichero ni las comillas)<sup>5</sup>.

```
> source("../script/funciones.R")
```

Comentaremos, brevemente, las funciones que han sido cargadas en esta primera versión y los parámetros de cómo han de ser usadas en todo momento.

Todas las funciones tienen el prefijo *isi*, seguido de guión bajo y el objetivo por la que fue creada. La primera función es *isi\_cargar*, como su nombre indica será la encargada de leer el o los archivos de texto y añadirlos a un *dataframe*. La forma de usar esta función es la siguiente:

```
WoS <- isi_cargar("../txt/")
```

Lo que le estamos diciendo a la función es que nos cargue en un *dataframe* llamado “WoS” todos los archivos que se encuentran en el directorio “../txt/” (no olvide ni las comillas ni el *slash* final). El usuario indicará la ruta donde tenga almacenado todos los .txt de su descarga. Puede nombrar el *dataframe* como desee, no es necesario que se llame WoS, lo usamos como ejemplo.

El proceso de carga suele ser un proceso que consume muchos recursos de nuestro ordenador y por tanto dependiendo del volumen de datos se puede volver muy lento.

Una vez que la función termina el proceso, tendremos en nuestro espacio de trabajo el *dataframe* con todos los datos cargados; ahora bien, para continuar disponemos de dos posibilidades: seguir trabajando con los datos cargados en la memoria o exportar los datos a una base de datos. Para esta segunda opción es necesario instalar en R ciertas librerías: dentro de la propia función se ha agregado verificación de que la librería se encuentra instalada y de no ser así, ésta la instala. Las librerías necesarias para continuar el proceso son: *plyr*; *DBI*; *RSQLite* y *ggplot2*.

Para exportar los datos a un sistema de bases de datos relacionales emplearemos la función ‘*isi\_exportar\_sqlite*’. Como su nombre indica, exporta el *dataframe* que le indiquemos a una base de datos *SQLite*. Esta función trabaja con dos parámetros, el primero ha de ser el nombre del *dataframe* que queremos exportar y el segundo la tabla donde queremos almacenar el contenido. La función se encuentra preparada para exportar a tablas los siguientes datos: el *dataframe* principal almacenado en la tabla que hemos llamado ‘*isi\_t\_pg*’; los autores en la tabla ‘*isi\_t\_au*’; las direcciones en la tabla ‘*isi\_t\_c1*’; las categorías temáticas en la tabla ‘*isi\_t\_wc*’ y finalmente las referencias bibliográficas en la tabla ‘*isi\_t\_cr*’. La base de datos la hemos llamado ‘*pgwos.db*’ y se creará en la raíz del disco o en la carpeta del proyecto.

Para exportar el *dataframe* WoS a la tabla principal de la base de datos ejecutamos la siguiente instrucción:

```
isi_exportar_sqlite(WoS,"WoS")
```

Si el usuario opta por continuar sin guardar el *dataframe* principal (**WoS**) en una BD, este podrá comenzar a obtener indicadores de los denominados unidimensionales, tales como frecuencias de datos referidos a la fecha, tipología documental, entre otros. Por ejemplo, una sencilla tabla de distribución de frecuencias de la variable fecha en el *dataframe* principal (**WoS\$PY**), podría obtenerse con las siguientes instrucciones:

```
> q=as.data.frame(table(WoS$PY))
> q=cbind(q,F.Acum=cumsum(q$Freq))
```

Cuyo resultado se presenta de la siguiente forma:

PY	Freq	F.Acum
2004	9	9
2005	16	25
2006	14	39
2007	22	61
2008	25	86
2009	26	112
2010	35	147
2011	31	178
2012	40	218
2013	18	236

La instrucción **table** genera una tabla de frecuencias (**PY**, **Freq**) de la variable elegida que se almacena en el *dataframe* (**q**). La instrucción **cbind** añade columnas a este *dataframe* y lo renombra igualmente (**q**), el argumento **cumsum** calcula la frecuencia acumulada de la variable (**PY**). Se anima al lector a probar con otras variables de ocurrencia simple -un único dato por registro- como las mencionadas más arriba o, por ejemplo, la referida al título de la fuente de publicación. Sin embargo, para realizar otros tipos de estudios, basados en campos de ocurrencias múltiples, es necesario procesar dichos datos separándolos en otros subconjuntos de datos. Este es el caso de los listados de autores, de direcciones, y categorías temáticas, entre otras. A continuación describimos el procedimiento para tal fin.

Crearemos subconjuntos de datos referentes a los autores, las direcciones, las categorías y las referencias bibliográficas (si éstas se han incluido en las descargas de datos originales). Para ello, es necesario emplear la función que hemos denominado

'isi\_separar', que admite como argumento alguno de los siguientes: 'C1' para crear el *dataframe* con las direcciones; 'AU' para los autores; 'WC' o 'SC' para las categorías temáticas y 'CR' para las referencias bibliográficas.

A continuación crearemos cada uno de los *dataframes* y luego los almacenamos en sus correspondientes tablas en la BD.

```
#Instrucciones que crean los diferentes Dataframes
autores <- isi_arar(WoS,"AU")
direcciones <- isi_separar(WoS,"C1")
temas <- isi_separar(WoS,"WC")
categorias <- isi_separar(WoS,"SC")
referencias <- isi_separar(WoS,"CR")
#Almacenamos en la BD
isi_exportar_sqlite(autores,"AU")
isi_exportar_sqlite(direcciones,"C1")
isi_exportar_sqlite(temas,"WC")
isi_exportar_sqlite(categorias,"SC")
isi_exportar_sqlite(referencias,"CR")
```

Para dar por concluida esta primera parte, se explica brevemente cómo almacenar los datos en otra fuente externa, para su posterior tratamiento, como puede ser una hoja de cálculo.

```
write.table(WoS, "wos.txt", sep="\t")
```

La instrucción anterior salva los datos en un archivo '.txt' separado por tabulaciones. Aunque también podemos usar:

```
write.csv(WoS, file = "wos.csv")
```

Si queremos salvar los datos en una hoja de Excel es necesario instalar y cargar la librería "xlsx"

```
library(xlsx)
write.xlsx(WoS, "wos.xlsx")
```

Adicionalmente podemos instalar y usar la librería "foreign" que permite importar y exportar datos a SPSS, SAS o Stata.

```
library(foreign)
write.foreign(WoS, "wos.txt", "wos.sps",
package="SPSS")
```

### 3.2. Obtención de indicadores

En este apartado mostramos algunas de las posibilidades de la metodología que aquí proponemos. Se ha programado, a modo de ejemplo, una función, 'isi\_unimetrics\_wos', que emplea la potente librería gráfica 'ggplot2' para obtener directamente algunos indicadores unidimensionales gráficos desde el *dataframe* principal (WoS).

La función "isi\_unimetrics\_wos" acepta dos parámetros para su funcionamiento, el primer parámetro es el nombre del *dataframe* principal y el se-

gundo argumento es el campo objeto del análisis. Se pueden analizar, en esta versión de la función, el campo PY referente a la fecha de publicación de los trabajos; el campo DT que muestra la tipología documental; LA para los idiomas. Se ha añadido un parámetro adicional, ACUM, para tener una tabla de frecuencias por fecha y el crecimiento acumulado. De este modo, la sintaxis `isi_unimetrics_wos(WoS,"PY")` obtiene como resultado un gráfico como el de la figura 1.

Para mostrar la flexibilidad de nuestra propuesta, se ha programado una función más, denominada 'h\_index' que calcula el *índice-h* de la distribución en un año determinado a partir de los datos almacenados en el *dataframe* WoS. La sintaxis de ejecución es:

```
h_index(WoS,"PY","2010")
```

```
## =====
## El índice h para el 2010 es: 7
## =====
```

La función admite un intervalo de fechas (i), para lo cual hay que ejecutarla con la siguiente sintaxis:

```
for(i in 2010:2014){h_index(WoS,"PY", i)}
```

Si se quiere calcular el *índice-h* en una determinada disciplina, utilizaremos la función del R-core, 'subset', con la siguiente sintaxis:

```
sub_WoS=subset(WoS, grepl("Physics*",WoS$WC))
```

el nuevo *dataframe* (sub\_WoS) podrá usarse con

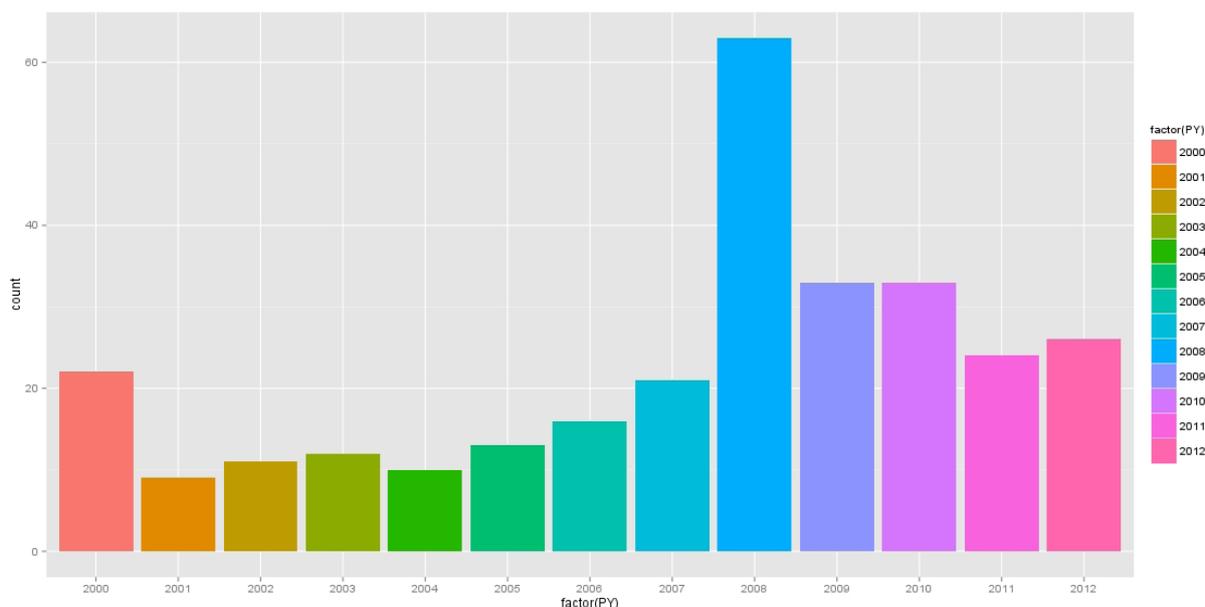
la función "h\_index", para calcular el *índice-h* (en un año o en un periodo). El argumento **grepl** del comando **subset**, permite aplicar expresiones regulares para recuperar el área temática correspondiente en el campo WC del *dataframe* principal (WoS).

#### 4. CONCLUSIONES

Sabemos que hay soluciones técnicamente mejores y más elegantes en su planteamiento y ejecución, pero lo que se persigue en este trabajo es presentar una colección de recursos asequibles y con los menores requerimientos informáticos y estadísticos posibles. Recursos, que no solo estén disponibles libremente, condición que nos autoimponemos, sino que sean fácilmente utilizables por quienes necesiten llevar a cabo estudios bibliométricos rigurosos. En particular hemos pensado que será útil a aquellos bibliotecarios que vayan asumiendo las nuevas funciones de apoyo y evaluación de la investigación. Complementariamente la metodología que se expone en este trabajo puede ser de interés para aquellos estudiantes de Información y Documentación que vayan a defender su Trabajo Final de Grado (TFG) en el ámbito de los estudios métricos de información.

Se presentan herramientas autónomas, que funcionan sin apenas control del usuario, pero la filosofía que se persigue es la de motivar a los lectores en la exploración de opciones más avanzadas tratando de convencer que con un mínimo esfuerzo, y desde luego una mayor motivación, pueden ad-

Figura 1.



quirirse las competencias necesarias para realizar análisis bibliométricos y cuantitativos fiables.

En la obtención de datos susceptibles de uso bibliométrico el formato `.csv` se ha convertido en un estándar *de facto*, con muchas alternativas de tratamiento. Las hojas de cálculo lo han sido, y en gran parte lo siguen siendo, pero aunque han mejorado en sus limitaciones globales, como el número de filas o columnas a tratar, aún siguen teniendo grandes limitaciones, como el número de caracteres que se admite en las celdas (255), y si tenemos valores en los campos que sobrepasen ese número no serán tratados correctamente, pensemos en algo cada vez más habitual como los trabajos *hiper-autorados* con centenares e incluso miles de autores y sus correspondientes direcciones institucionales.

El usar una base de datos relacional en este método nos aporta las bondades del lenguaje SQL,

con lo que podemos organizar los datos de una forma más eficiente y con un mayor potencial en su tratamiento que lo que puede resultar del uso de gestores bibliográficos o de las herramientas de análisis incorporadas en las fuentes bibliográficas utilizadas. Hay vías que sólo se insinúan en este trabajo, como es la posibilidad de la consulta directa, sin cliente o gestor de base de datos, desde herramientas globales como es el propio software R.

Por último, deseamos hacer notar que lo que proponemos en este trabajo es una adaptación a los nuevos tiempos de la vieja filosofía del *'desktop scientometrics'*, esta vez basada en aplicaciones multiplataforma y de Software Libre que permita convencer a los usuarios de que con pequeños esfuerzos y mentalidades abiertas al cambio tecnológico es posible disponer en la palma de su mano de herramientas muy potentes para elaborar una amplia variedad de indicadores de interés bibliométrico.

## 5. NOTAS

- [1] La Universidad de Leicester fue la primera en el Reino Unido en contratar a un especialista en bibliometría (<http://www2.le.ac.uk/news/blog/2010-archive/march-2010/first-uk-academic-bibliometrician-appointed>).
- [2] Por su utilidad recomendamos el entorno de trabajo RStudio™ (<http://www.rstudio.com/>) que es una interfaz integrada libre (no la única) para trabajar con R de una manera más cómoda y amigable.
- [3] Disponible en <https://addons.mozilla.org/es/firefox/addon/sqlite-manager/>.

- [4] Además del *script* con las funciones que se presentan en este trabajo ("funciones.r"), está disponible a modo de guía práctica una versión reducida de este trabajo ("README.md").
- [5] Si usa el *IDE RStudio*, lo más recomendable es iniciar un proyecto nuevo y crear dentro de él una estructura de carpetas. En nuestro caso el fichero está en una carpeta denominada `/script/`, y los ficheros descargados de la WoS en otro denominado `/txt/`. Si se usa esta opción las rutas deberán comenzar por `../`.

## 6. REFERENCIAS

- Anuradha, K.T.; Urs, S.R. (2007). Bibliometric indicators of Indian research collaboration patterns: A correspondence analysis. *Scientometrics*, vol. 71, no. 2, pp. 179-189. <http://dx.doi.org/10.1007/s11192-007-1657-4>
- Arriaza Gómez, A. J.; Fernández Palacín, F.; López Sánchez, M. A.; Muñoz Márquez, M.; Pérez Plaza, S.; Sánchez Navas, A. (2008). *Estadística básica con R y R-Commander* [en línea]. Cádiz: Universidad de Cádiz, Servicio de Publicaciones. [Consulta: 10 enero 2015]. Disponible en: <http://gsyc.escet.urjc.es/~herreraiz/ebrcmdr.pdf>
- Bowman, T.D.; Tsou, A., Ni; C.; Sugimoto, C.R. (2014). Post-interdisciplinary frames of reference: exploring permeability and perceptions of disciplinarity in the social sciences. *Scientometrics*, vol. 101, no. 3, pp. 1695-1714. <http://dx.doi.org/10.1007/s11192-014-1338-z>
- Carlson, J.; Kneale, R. (2011). Embedded librarianship in the research context: Navigating new waters. *College & Research Libraries News*, vol. 72, no. 3, pp. 167-170.
- Chen, H.; Zhao, G.; Xu, N. (2012). The Analysis of Research Hotspots and Fronts of Knowledge Visualization Based on CiteSpace II. En: S.K.S. Cheung, J. Fong, L.-F. Kwok, K. Li y R. Kwan (eds.), *Hybrid Learning* [en línea]. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 57-68. [Consulta: 10 enero 2015]. Disponible en: [http://link.springer.com/10.1007/978-3-642-32018-7\\_6](http://link.springer.com/10.1007/978-3-642-32018-7_6).
- Corrall, S.; Kennan, M.A.; Afzal, W. (2013). Bibliometrics and Research Data Management Services: Emerging Trends in Library Support for Research. *Library Trends*, vol. 61, no. 3, pp. 636-674. <http://dx.doi.org/10.1353/lib.2013.0005>
- Dewey, B.I. (2004). The Embedded Librarian. *Resource Sharing & Information Networks*, vol. 17, no. 1-2, pp. 5-17. [http://dx.doi.org/10.1300/J121v17n01\\_02](http://dx.doi.org/10.1300/J121v17n01_02)

- Gagolewski, M. (2011). Bibliometric impact assessment with R and the CITAN package. *Journal of Informetrics*, vol. 5, no. 4, pp. 678-692. <http://dx.doi.org/10.1016/j.joi.2011.06.006>
- García González, P.E. (2010). *Diseño, desarrollo y aplicación de un método para el análisis y tratamiento de la información con fines métricos*. [Tesis doctoral]. Universidad Carlos III de Madrid.
- Grauwin, S.; Jensen, P. (2011). Mapping scientific institutions. *Scientometrics*, vol. 89, no. 3, pp. 943-954. <http://dx.doi.org/10.1007/s11192-011-0482-y>
- Han, G.; Sun, B. (2014). Design and Implementation of Bibliometrics System Based on RIA. En: S. Patnaik, S.; Li, X. (editores), *Proceedings of International Conference on Soft Computing Techniques and Engineering Application* [en línea]. New Delhi: Springer India, pp. 263-272. [Consulta: 10 enero 2015]. Disponible en: [http://link.springer.com/10.1007/978-81-322-1695-7\\_30](http://link.springer.com/10.1007/978-81-322-1695-7_30)
- Hernández-Pérez, T.; Pacios, A.R.; Vianello, M.; Ortega, R.; Gorospe, M.R. (2011). La formación en alfabetización en información en las aulas universitarias: el caso de la UC3M. *Scire: representación y organización del conocimiento*, vol. 17, no. 2, pp. 27-37.
- Katz, J.; Hicks, D. (1997). Desktop Scientometrics. *Scientometrics*, vol. 38, no. 1, pp. 141-153. <http://dx.doi.org/10.1007/BF02461128>
- Kesselman, M.A.; Watstein, S.B. (2009). Creating Opportunities: Embedded Librarians. *Journal of Library Administration*, vol. 49, no. 4, pp. 383-400. <http://dx.doi.org/10.1080/01930820902832538>
- Leydesdorff, L.; Rafols, I. (2012). Interactive overlays: A new method for generating global journal maps from Web-of-Science data. *Journal of Informetrics*, vol. 6, no. 2, pp. 318-332. <http://dx.doi.org/10.1016/j.joi.2011.11.003>
- Paradis, E. (2005). *R for Beginners* [en línea]. [Consulta: 10 enero 2015]. Disponible en: [http://cran.r-project.org/doc/contrib/rdebuts\\_es.pdf](http://cran.r-project.org/doc/contrib/rdebuts_es.pdf)
- R Core Team. (2012). *R: A Language and Environment for Statistical Computing* [en línea]. Vienna, Austria: R Foundation for Statistical Computing. [Consulta: 10 enero 2015]. Disponible en: <http://www.R-project.org>
- Zhang, L.; Thijs, B.; Glänzel, W. (2013). What does scientometrics share with other «metrics» sciences? *Journal of the American Society for Information Science and Technology*, vol. 64, no. 7, pp. 1515-1518. <http://dx.doi.org/10.1002/asi.22834>