

ESCALONA FERNÁNDEZ, M.ªI., LAGAR BARBOSA, P. y PULGARÍN GUERRERO, A. Web of Science vs. SCOPUS: un estudio cuantitativo en Ingeniería Química. *Anales de Documentación*, 2010, vol. 13, p. 159-175.

WEB OF SCIENCE VS. SCOPUS: UN ESTUDIO CUANTITATIVO EN INGENIERÍA QUÍMICA

María Isabel Escalona Fernández

Servicio de Biblioteca, Archivo y Documentación. Universidad de Extremadura.

Pilar Lagar Barbosa

Departamento de Información y Comunicación. Universidad de Extremadura.

*Antonio Pulgarín Guerrero**

Departamento de Información y Comunicación. Universidad de Extremadura.

Resumen: El propósito de este estudio es ofrecer a los investigadores químicos, una visión clara y práctica, aplicando métodos cuantitativos, sobre las dos plataformas de información científica más extendidas hoy día: Web of Science (Thomson Reuters) y Scopus (Elsevier). Se llevaron a cabo, en ambas bases de datos una búsqueda especializada en el área de la ingeniería química, entre 1999 y 2008. A través de los artículos recuperados en las búsquedas, durante esos diez años, y las revistas recuperadas de 2006, se analiza la posible correlación entre ambos sistemas; el modelo de crecimiento que presentan y ciertos parámetros de ese crecimiento; el solapamiento entre las dos bases de datos, la dispersión de los artículos en las revistas y medidas de concentración, entre otros. Los resultados muestran la existencia de una alta semejanza entre Web of Science y Scopus, resultando complementarios, pero no excluyentes, respecto a su posible utilización por los ingenieros químicos.

Palabras clave: Web of Science; Scopus; métodos cuantitativos; crecimiento; solapamiento; dispersión; concentración; informetría.

Title: WEB OF SCIENCE VS. SCOPUS: A QUANTITATIVE STUDY IN CHEMICAL ENGINEERING.

Abstract: The objective of this study is to offer a clear and practical overview to the Chemical Scientist, applying quantitative methods on the two most extended platforms of scientific information nowadays: Web of Science (Thomson Reuters) and Scopus (Elsevier). It was carried out a specialized search in the area of the chemical engineering in both databases between 1999 and 2006. Through the articles recovered in the searches during those ten years and the journals recovered in 2006, it is analyzed the possible correlation between both systems; the pattern of growth that they show and certain parameters of that growth; the overlapping among the two databases, the dispersion of the articles in the journals and concentration measures, between others. The results show the existence of a high likeness between Web of Science and Scopus, turning out complementary but not exclusive, regarding their possible use for the chemical engineers.

Keywords: Web of Science; Scopus; quantitative methods; growth; overlapping; dispersion; concentration; informetrics.

* pulgarin@unex.es

1. INTRODUCCIÓN

1.1 Justificación y objetivo

La idea del presente trabajo surge como consecuencia de otra publicación de los autores en 2007, sobre solapamiento de tres bases de datos, en ingeniería (Pulgarín y Escalona, 2007).

La ingeniería química es una de las áreas de conocimiento que se pueden considerar ausentes de sesgo, respecto al tratamiento de los dos sistemas de bases de datos que aquí se estudian: Web of Science (Thomson Reuters) y Scopus (Elsevier).

En la literatura más reciente han sido numerosos los estudios llevados a cabo sobre estos dos sistemas de bases de datos, si bien se refieren, en su mayoría, a aspectos generales, como sistemas de búsqueda, facilidades y características que presentan cada uno de ellos, etc. Son artículos con descripciones y opiniones sobre los interfaces, capacidades y caracteres que presentan cada uno de estos productos, dándose a entender, por la literatura consultada, que hace una década “Web of Science” (WoS) de Thomson Reuters no tenía contrincante, pero que de unos años acá, se le ha presentado un excelente adversario, “Scopus” de Elsevier. Esta es la conclusión general que se podría obtener de estos estudios. Pero un estudio, basado en el contenido de estas dos bases de datos y empleando herramientas del análisis matemático y estadístico, en una materia como la ingeniería química, no se había realizado hasta ahora.

1.2 Breve revisión bibliográfica sobre comparación entre WoS y Scopus

Entre los trabajos anteriores, sobre la comparación entre WoS y Scopus, destacan una revisión de Goodman y Deis (2005) y la posterior actualización por estos mismo autores (Goodman y Deis, 2007). La revisión trata de comparar la versión 7 del WoS y Scopus. Ambos productos son bases de datos académicas con contenido en materia multidisciplinar.

En la revisión de 2005, Goodman y Deis describen y comparan los precios, los productos que ofrecen ambos sistemas, cobertura respecto a contenido y tiempo, actualización, tipos de documentos que vacía, facilidades de búsquedas, distribución de documentos, acceso, etc. Finalmente, ofrecen una serie de recomendaciones a tener en cuenta.

En la actualización de 2007, los autores ponen al día los datos, los nuevos contenidos y como conclusión y recomendaciones no ofrecen una clara elección por uno u otro sistema. Al existir, según los autores, mínimas diferencias no toman ninguna decisión al respecto. Mientras que Scopus ofrece una amplia selección de revistas y un interface inteligente, WoS ofrece una mayor cobertura en años. Por tanto serían complementarios, dicen los autores.

Entre los dos trabajos de Goodman y Deis han aparecido otros con el mismo objetivo, comparar las características entre Scopus y WoS.

Uno de ellos como homenaje a Garfield, cuando se cumplen 50 años de la creación del índice de citas, donde el autor Peter Jacso, hace un símil entre este hecho y el producido hace 60 años por Vannevar Bush, al inventarse una estación de trabajo de información conocida como Memex (Jacso, 2005).

Otros dos trabajos, comparando ambas bases de datos, han sido publicados por LaGuardia (2005) y Fingerman (2006).

En 2009, Archambault *et al.* han publicado los resultados de un análisis comparativo del ranking de países por número de publicaciones y número de citas recibidas, obteniendo una alta correlación entre WoS y Scopus.

1.3 Revisión sobre solapamiento de fuentes bibliográficas

Las primeras investigaciones sobre solapamiento de revistas y/o de fuentes secundarias tuvieron lugar en los pasados años 60. Martyn y Slater (1964), fueron de los primeros autores que trabajaron sobre este asunto, si bien hubo otros menos importantes, que o bien no cubrieron suficientes áreas de conocimiento o no presentaron un marco teórico.

Los trabajos de Martín (1967) esbozan una fácil teoría sobre el solapamiento de revistas y presentan varios estudios de casos de fuentes secundarias. Pero, tanto en el análisis metodológico como en la selección de materiales no fueron lo suficientemente robustos, desde un punto de vista estadístico, y muchas de sus conclusiones fueron invalidadas (Bost, 1968).

Otros autores de los 60, simplificaron el análisis del solapamiento, usando los títulos de las fuentes primarias e ignorando el análisis del solapamiento respecto a los artículos en las revistas de resúmenes. Entre ellos destaca el análisis efectuado por Bourne *et al.* (1969). Se trata de una extensiva revisión de la relación entre la *Bibliography of Agricultura (BoFA)* con otras 15 fuentes secundarias.

El estudio más extenso sobre fuentes secundarias científico-técnicas fue llevado a cabo, entre 1974-1977, por la National Science Foundation's Office of Science Information Service, de USA, (Berman y Kunberger, 1977). El estudio abarca el solapamiento entre 14 grandes fuentes, entre ellas el *Chemical Abstracts*, *Biological Abstracts*, e *Engineering Index*. Este trabajo supone uno de los más importantes análisis de solapamiento de fuentes secundarias: la autora introduce la necesidad de usarlo de forma más amplia, con medidas menos sencillas de solapamiento y sugiere varias definiciones alternativas para el solapamiento. Utiliza herramientas matemáticas más sofisticadas para el análisis, empleando el concepto de doble forma respecto al solapamiento de las fuentes secundarias.

Bearman fue, también, la primera en aplicar técnicas estadísticas, como análisis factorial y escalado multidimensional (MDS) al análisis de solapamiento de datos, identificando varios grupos naturales como física, ciencias de la tierra o ciencias de la vida. Igualmente investigó varias fórmulas para medir la proximidad de fuentes secundarias usando MDS.

Después del estudio de Bearman, otros estudios más restringidos se han llevado a cabo, incluyendo, la mayoría de ellos, análisis de solapamiento múltiple. Algunos de estos estudios posteriores han eliminado muchos de los análisis estadísticos y de muestreos: usando la población total y estableciendo los actuales valores del solapamiento tradicional (Poyer, 1984).

En 1990 aparece una revisión sobre solapamiento, en la que su autor (Gluck, 1990) revisa las investigaciones sobre solapamiento realizadas en las últimas décadas e introduce una extensión a la tradicional definición de solapamiento. La extensión consiste en la incorporación, a la ya tradicional definición, de los tamaños relativos de las fuentes

en una matriz de valores de desemejanza. Además aplica técnicas de escalado multidimensional (MDS) para demostrar, gráficamente, este concepto modificado de solapamiento. Según Gluck, la definición tradicional de solapamiento ignora el tamaño de las fuentes.

Para Gluck, TO es una medida de semejanza, en la jerga del MDS. No obstante, el concepto dual de separación o de libertad de dos fuentes representa una medida de desemejanza. La extensión que propone explota el potencial de la desemejanza, al ser más manejable que la semejanza para ciertos algoritmos de MDS, para incrementar el análisis estadístico, convirtiendo la medida tradicional en una medida de desemejanza. Por ejemplo, la medida de semejanza, TO, se puede convertir en otra de desemejanza, tan solo restando su valor a la unidad.

Esta medida, propuesta Gluck, contabiliza el tamaño relativo de las distintas fuentes y el cambio de dirección de la tradicional definición de solapamiento.

Finalmente, Gluck introduce lo que él llama 'Syndetic freedom', SF, como alternativa a TO, e ilustra el proceso comparando SF y TO, con la base de datos CAB Abstracts, concretamente utiliza tres ficheros de CAB y resultados de tres años de estos ficheros.

Más recientemente, Hood y Wilson (2003), en un estudio del solapamiento sobre "teoría de conjuntos borrosos" en una serie de bases de datos, han encontrado que un 63% de los registros aparecieron en una sola base de datos, mientras que el 37% restante estaban duplicados entre 2 y 12 bases de datos.

Y los autores del presente trabajo, publican en 2007 un estudio sobre el solapamiento entre las bases IEL, INSPEC y PASCAL, con información en ingeniería (Pulgarín y Escalona, 2007).

2. METODOLOGÍA

Para realizar el presente trabajo se han llevado a cabo una serie de búsquedas retrospectivas, sobre ingeniería química, tanto en WoS como en Scopus, por separado, que abarcan el período de tiempo de 1999 a 2008.

Igualmente, se han recuperado todos los títulos de revistas de lo publicado en 2006, en ambos sistemas, también por separado.

Una dificultad importante, con la que los autores se han encontrado fue, precisamente, efectuar las búsquedas, en WoS y Scopus, de la forma más similar posible, ya que el lenguaje de interrogación de cada sistema es distinto. Equiparar y verificar las sensibilidades de los dos sistemas fue la tarea que mayor dificultad presentó, eso es, obtener unos resultados lo menos sesgados posible, como consecuencia de las búsquedas realizadas, independientemente de la política editorial o de vaciado de las revistas que tenga cada sistema.

Las ecuaciones de búsquedas ejecutadas, tanto en WoS como en Scopus han sido las siguientes:

Búsqueda en WoS

Year Published=(2007) AND Address=("ing* quim*" or "chem* eng*")
Refined by: Document Type=(ARTICLE OR PROCEEDINGS PAPER)
Timespan=All Years. Databases=SCI-EXPANDED, SSCI, A&HCI.

Búsqueda en Scopus

AFFIL((ing* PRE/0 quim*) OR (chem* PRE/0 eng*)) AND PUBYEAR IS 2007 AND DOCTYPE(ar OR cp) AND SRCTYPE(j OR d).

Para efectuar el análisis se utilizaron diversas herramientas matemáticas y estadísticas, para el tratamiento de los datos y el cálculo de ciertos estadígrafos y parámetros, relacionados con el crecimiento de la literatura científica, en el período de tiempo considerado.

La existencia o no de correlación, respecto a la producción anual, entre las dos bases de datos, fue determinada mediante el cálculo del coeficiente de correlación y/o determinación. Para comprobar la diferencia entre las medias de ambas muestras se utilizó la “*t*” de Student y para comprobar la homogeneidad de las varianzas se aplicó el estadígrafo “*F*” de Snedecor. La finalidad de los dos test fue comprobar si las muestras de WoS y de Scopus, procedían de la misma población.

Los parámetros calculados se refieren a la tasa de crecimiento, el número de artículos producido por unidad de tiempo, el tiempo de duplicación de la literatura científica de ingeniería química y un pronóstico a 5 años, teniendo en cuenta el modelo exponencial de Price (1963) (ver Anexo).

Con objeto de comprobar el grado de solapamiento entre WoS y Scopus, se llevan a cabo una serie de cálculos basados en las fuentes primarias (revistas) que cubren ambas fuentes secundarias (las bases de datos). El problema radica en que mientras unos productores de bases de datos vacían las fuentes completamente, de cubierta a cubierta, otros lo hacen de forma selectiva (Pao, 1993).

Finalmente, al conjunto de revistas, correspondientes al año 2006, en ambas bases de datos por separado, se les aplicaron medidas de concentración (curva de Lorenz e Índice de Gini) y se observó la dispersión de los artículos en las mismas. Se seleccionó el año 2006 por ser de los más recientes y al mismo tiempo que garantizase la cobertura total, esto es, que estuviesen vaciadas todas las revistas correspondiente a ese año, ya que a veces existen retrasos.

3. RESULTADOS

3.1 Crecimiento

En WoS, el crecimiento del número de artículos, sobre ingeniería química, fue de 10.368, en 1999, a 22.577, en 2008; y en Scopus, de 10.352, en 1999, a 24.780, en 2008.

La distribución anual de estos resultados se muestra en la Tabla I y la tendencia ascendente, en ambos casos, se puede observar en la Figura 1.

Años	WoS	Scopus
1999	10368	10352
2000	11030	11040
2001	11924	12043
2002	13046	12435
2003	14193	14408
2004	15568	16742
2005	17088	18937
2006	19169	21355
2007	20997	23193
2008	22577	24780
Total	155960	165285

Tabla I. Evolución de la producción científica sobre ingeniería química en WoS y Scopus.

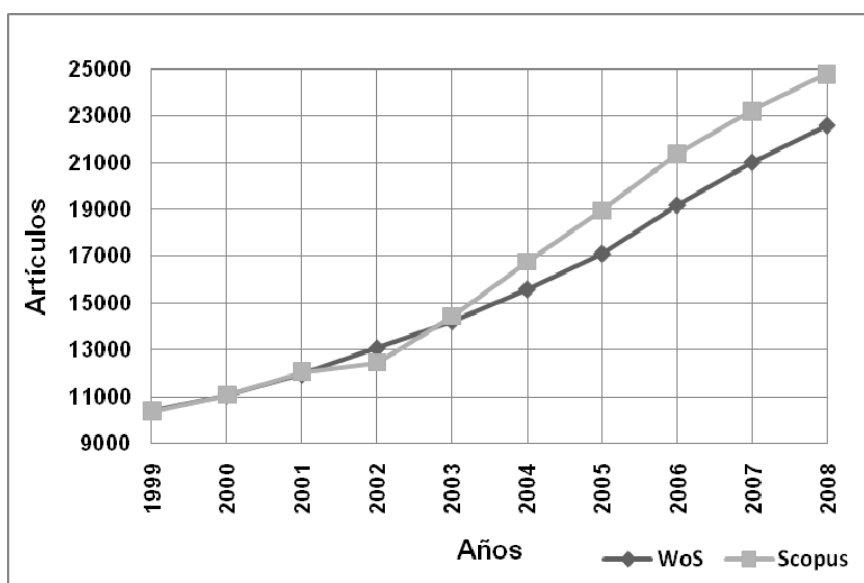


Figura 1. Evolución del número de artículos sobre ingeniería química, en WoS y en Scopus.

El crecimiento experimentado, en ambos casos, ha sido exponencial, en el período de tiempo estudiado (1999-2008), con un coeficiente de determinación $R^2 = 0,996$, para WoS y de $0,985$, para Scopus.

Al disponer la distribución anual del número de artículos de WoS frente al de Scopus, se comprueba la existencia de una fuerte correlación entre ambos. La Figura 2, muestra la representación de estos datos y su ajuste a una recta, con un coeficiente de correlación igual a $0,996$ ($R^2 = 0,992$).

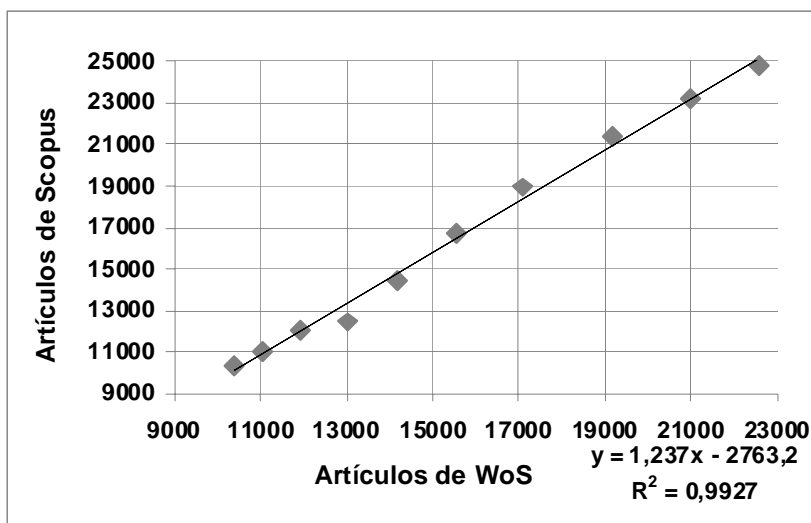


Figura 2. Correlación entre WoS y Scopus.

3.1.1 Diferencia de medias

Si suponemos que $\sigma_1 = \sigma_2 = \sigma$, siendo σ desconocida, cuando el tamaño de las muestras es pequeño ($n, m \leq 30$), la diferencia de medias muestrales (\bar{x}, \bar{y}) sigue una distribución t de Student, con $n+m-2$ grados de libertad, de la forma siguiente:

$$t_{n+m-2} = \frac{\bar{x} - \bar{y}}{s_{\bar{x}-\bar{y}}}, \tag{1}$$

El error estándar estimado de la diferencia entre medias, cuando las muestras tiene el mismo tamaño, se calcula como

$$s_{\bar{x}-\bar{y}} = \sqrt{\frac{SC_n + SC_m}{n(n-1)}}, \tag{2}$$

SC_n , es la suma de los cuadrados de la muestra n .

$$SC_n = \sum x_i^2 - \frac{(\sum x_i)^2}{n}$$

SC_m , es la suma de los cuadrados de la muestra m .

$$SC_m = \sum y_i^2 - \frac{(\sum y_i)^2}{m}$$

El error estándar estimado de la diferencia de medias fue de 2.145,6 y, por lo tanto, el estadígrafo t resultó ser 0,435.

$$t_{n+m-2} = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{SC_n + SC_m}{n(n-1)}}} = \frac{16.528,5 - 15.596}{2.145,6} = \frac{932,5}{2.145,6} = 0,435.$$

Como el valor crítico de la t , a un nivel de significación de $\alpha = 0,05$ y 18 grados de libertad, es igual a 2,1 (región crítica), valor que es superior que el de la t obtenida (0,435), se acepta la hipótesis nula (H_0 : “no hay diferencia significativa entre las medias de las dos muestras”), a un nivel de significación del 5%.

3.1.2 Homogeneidad de las varianzas

Aún no podemos afirmar que las dos muestras (m , n) procedan de la misma población, a pesar de no ser significativamente diferentes.

Para ello, se deberán estimar las varianzas y determinar si difieren o no. El estadígrafo utilizado para llevar a cabo el contraste de la homogeneidad de las varianzas es la F de Snedecor, con $m-1$ y $n-1$ grados de libertad.

El cociente F se define de la forma siguiente:

$$F = \frac{\hat{S}_m}{\hat{S}_n}; \left(\hat{S}_m > \hat{S}_n \right) \quad (3)$$

$$\hat{S}_m = \frac{SC_m}{N_m - 1} = \frac{251.290.127}{9} = 27.921.125,2$$

$$\hat{S}_n = \frac{SC_n}{N_n - 1} = \frac{163.033.172}{9} = 18.114.796,9$$

Y el contraste de la F es:

$$F = \frac{\hat{S}_m}{\hat{S}_n} = \frac{27.921.152,2}{18.114.796,9} = 1,54$$

Como el valor crítico de la F a un nivel de significación de $\alpha = 0,05$ (bilateral) y 9/9 grados de libertad es igual a 3,18, valor superior que el de la F obtenida (1,54), se acepta la hipótesis nula (H_0 : “no hay diferencia significativa entre las varianzas”), a un nivel de significación del 5 % bilateral.

Con estas dos pruebas estadísticas se puede afirmar que ambas muestras proceden de la misma población, con una posibilidad de error del 5%.

3.1.3 Cálculo de la tasa de crecimiento (ver Anexo)

Para el caso de WoS, el número de artículos en 1999 fue de 10.368 y de 22.577 en 2008. Para Scopus, el número de artículos recuperados en 1999 fue de 10.352 y de 24.780 en 2008.

La tasa de crecimiento se calculó utilizando la fórmula:

$$P(t) = P(0) \cdot e^{a \cdot t} \quad (4)$$

La tasa obtenida para WoS fue:

$$22577 = 10368 \cdot e^{a \cdot 10}; \frac{22577}{10368} = e^{a \cdot 10}; 2,177 = e^{a \cdot 10}; \ln 2,177 = 10 \cdot a \cdot \ln e; 0,77 = 10 \cdot a; a = 0,077.$$

Y la tasa de crecimiento obtenida para Scopus fue: $a = 0,087$.

Se observa que ambas tasas son muy similares.

3.1.4 Cálculo del número de artículos por unidad de tiempo (ver Anexo)

El cálculo del número de artículos por unidad de tiempo (año) es la derivada de $P(0) \cdot e^{a \cdot t}$, esto es

$$P'(t) = P(0) \cdot a \cdot e^{a \cdot t}. \quad (5)$$

Para los datos de WoS tenemos que

$$P'(t) = 10.368 \cdot 0,077 \cdot e^{0,077 \times 10} \cong 1.724 \text{ artículos.}$$

Y para Scopus

$$P'(t) = 10.352 \cdot 0,087 \cdot e^{0,087 \times 10} \cong 2.149 \text{ artículos.}$$

3.1.5 Cálculo del tiempo de duplicación (ver Anexo)

El tiempo de duplicación para la ingeniería química, es el tiempo que tarda en convertirse $P(t)$ en $2 \cdot P(t)$:

Para WoS:

$$\frac{2P(t)}{P(t)} = e^{a \cdot t}; 2 = e^{0,077 \cdot t}; \ln 2 = 0,077 \cdot t \cdot \ln e; 0,6931 = 0,077 \cdot t; t \cong 9 \text{ años.}$$

En Scopus el tiempo de duplicación fue de $t = 8$ años.

Los resultados estarían algo por debajo de la idea expresada por Price (1963), en la que anuncia que la ciencia se duplica cada 10-15 años, en personal o en publicaciones.

3.1.6 Cálculo del pronóstico a 5 años (ver Anexo)

Para un pronóstico, en el que el intervalo de tiempo es pequeño, por ejemplo un año, se podría utilizar la fórmula:

$$P(t + \Delta t) = P(t) + P'(t) \cdot \Delta t. \quad (6)$$

Sin embargo, a medida que el intervalo crece, también aumentan las diferencias. Por lo tanto, para el tiempo que nos planteamos, 5 años, será más precisa la fórmula (4):

$$P(t) = P(0) \cdot e^{a \cdot t}; P(t) = 22.577 \cdot e^{0,077 \cdot 5} = 33.179.$$

Para WoS, para el año 2.011, se obtuvo un pronóstico de 33.179 artículos:

Mientras que con la fórmula (6):

$P(t + \Delta t) = P(t) + P'(t) \cdot \Delta t = 22.577 + (1.724 \times 5) = 31.197$ artículos, resultado algo menor que el anterior.

Para Scopus, los resultados fueron de 38.285 artículos, para el primer caso, y de 35.525 utilizando la segunda fórmula.

3.2 Solapamiento entre WoS y Scopus

El % de solapamiento entre WoS y Scopus, solapamiento tradicional, TO, definido por Gluck (1990), fue del 70 %:

$$\begin{aligned} \%TO &= 100 \cdot \left(\frac{\# \text{de } A \cap B}{\# \text{de } A \cup B} \right) = 100 \cdot \left(\frac{\# \text{ISI} \cap \text{SCOPUS}}{(\# \text{ISI} \cup \text{SCOPUS}) - (\# \text{ISI} \cap \text{SCOPUS})} \right) = \\ &100 \cdot \frac{1398}{(1514 + 1880) - (1398)} = 70\% \end{aligned} \quad (7)$$

Este resultado se interpreta como que entre WoS y Scopus hay un 70 % de semejanza o, como también definió Gluck, un 30 % de diferencia o de separación entre ambas bases de datos.

Gluck, también definió, como alternativa a la TO, el “*Syndetic freedom*”, *SF*, como:

$$SF(A, B) = \left(1 - \frac{INT}{UNI} \right) * \left(\frac{MIN}{MAX} \right), \text{ para cuando } A \not\subset B. \quad (8)$$

Donde A y B son dos fuentes secundarias, INT el resultado de la intersección de esas dos fuentes, UNI el de la unión, MIN el valor mínimo de las dos fuentes y MAX el valor máximo de las mismas.

Esta medida contabiliza el tamaño relativo de los dos sistemas.

$$\begin{aligned} SF(WoS, Scopus) &= \left(1 - \frac{(\# WoS \cap Scopus)}{(\# WoS \cup Scopus) - (\# WoS \cap Scopus)} \right) * \left(\frac{\# WoS}{\# Scopus} \right) = \\ &\left(1 - \frac{1398}{(1514 + 1880) - 1398} \right) \left(\frac{1514}{1880} \right) = (1 - 0.7) \times 0.085 = 0.24 \end{aligned}$$

Para medir el porcentaje de cobertura de WoS, respecto a Scopus y a la inversa, se utilizó el solapamiento relativo, propuesto por Bearman y Kunberger (1977), definido como:

$$\%TO \text{ de } WoS = 100 \cdot \left(\frac{\# \text{de } WoS \cap Scopus}{\# \text{de } WoS} \right) = 100 \cdot \frac{1398}{1514} = 92.7\%. \quad (9)$$

Lo que quiere decir que Scopus cubre el 92 % de WoS.

$$\%TO \text{ de } Scopus = 100 \cdot \left(\frac{\# \text{de } Scopus \cap WoS}{\# \text{de } Scopus} \right) = 100 \cdot \frac{1398}{1880} = 74.3\%.$$

O que WoS cubre el 74.3 % de Scopus, sobre ingeniería química, basándonos en los datos de 2006.

3.3 Medida de la concentración y dispersión de los artículos

La concentración de los artículos entre las revistas resultó bastante alta, ya que, en el caso de WoS (Figura 3), tan solo el 2 % de las revistas publicaron más del 31 % de los artículos, con el 10 % de las revistas se alcanza el 65 % de los artículos y con tan solo el

20 % de las revistas se llega a casi el 90 % de la productividad total en ingeniería química. El Índice de Gini obtenido fue de 0,8 (Lorenz, 1905).

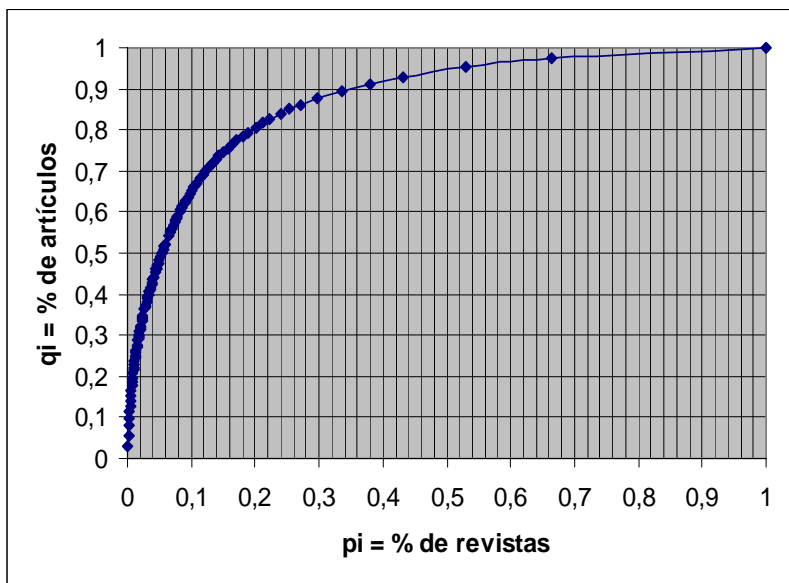


Figura 3. Curva de Lorenz (concentración) para WoS.

Similares resultados se obtuvieron para el caso de Scopus.

Respecto a la dispersión de los artículos, según los resultados de las búsquedas para el año 2006, fue también alta. Un total de 19.169 artículos, en el caso de WoS, fueron publicados en 1.514 revistas distintas (12,56 artículos/revista), mientras que para Scopus los 21.355 artículos recuperados fueron publicados en un total de 1.880 revistas (11,4 artículos/revista).

La tabla II muestra las revistas más productivas en ingeniería química, en el año 2006, para WoS y para Scopus. Las revistas, en número de 37 para WoS y 43 para Scopus, se disponen en orden decreciente de productividad de artículos. Este conjunto de revistas forman un núcleo cuya producción supera, en ambos casos, 1/3 de la producción total.

Nombre de la revista	WoS	Scopus
Acta Crystallographica Section E: Structure Reports Online	578	571
Industrial and Engineering Chemistry Research	485	476
Journal of Applied Polymer Science	431	421
Journal of Physical Chemistry B	364	353
Chemical Engineering Sciences	290	279
Journal of Membranes Science	265	258
Desalination	258	262
Langmuir: the ACS journal of surfaces and colloids	256	260
Journal of Power Sources	226	216

Nombre de la revista	WoS	Scopus
Macromolecules	215	193
AIChE Journal	190	182
Polymer	188	186
Journal of Chemical Industry and Engineering (China)	X	206
Journal of Colloid and Interface Science	179	176
Journal of Hazardous Materials	155	157
Applied Physics Letters	147	139
Journal of Chemical Physics	146	142
Journal American Chemical Society	139	128
Journal of Chemical and Engineering Data	135	132
Gongneng Cailiao/Journal of Functional Materials	X	119
Chinese Journal of Inorganic Chemistry	128	125
Xiandai Huagong/Modern Chemical Industry	X	124
Applied Catalysis A: General	121	122
Chinese Journal of Analytical Chemistry	116	60
Chemical Engineering (China)	X	121
Studies in Surface Science and Catalysis	X	119
Chemical Journal of Chinese Universities	112	111
Journal of the Electrochemical Society	111	107
Fluid Phase Equilibria	110	112
Biotechnology and Bioengineering	110	111
Colloids and Surfaces A: Physicochemical and Engineering Aspects	104	105
Acta Chimica Sinica	102	X
Journal of Chemical Engineering of Japan	101	94
Korean Journal of Chemical Engineering	100	98
Electrochimica Acta	100	98
Applied Surface Science (1985)	96	92
Separation and Purification Technology	94	91
Spectroscopy and Spectral Analysis	92	94
Biochemical Engineering Journal	92	87
Journal of Polymer Science, Part A: Polymer Chemistry	91	91
Catalysis Today	89	86
Chemical Engineering Journal	84	93
Petrochemical Technology	X	92
Journal of Chemical Engineering of Chinese Universities	X	90

Tabla II. Ranking de las revistas más productivas, en WoS y en Scopus.

4. CONCLUSIONES

Al igual que concluyeran Goodman y Deis (2005, 2007), aunque en este estudio se haya empleado otra metodología, el conjunto de conclusiones obtenidas en el mismo se podrían resumir en una sola; que los dos sistemas son muy similares tanto en contenido

como en dinámica, siendo complementarios de cara a su utilización por parte de los ingenieros químicos.

Esta similitud se ha podido comprobar comparando la producción científica, año a año, en ambos sistemas, así como el número de revistas indexadas en el año 2006.

La producción de artículos se ha duplicado holgadamente, en los diez años considerados, en ambas bases de datos (tiempo de duplicación para WoS de 9 años y de 8 para Scopus). La tasa de crecimiento fue de 7,7% para WoS y del 8,7 para Scopus, lo que supuso un crecimiento de más de 1.700 artículos en WoS y de 2.100 en Scopus, por año.

La tendencia claramente ascendente de la producción científica presenta un modelo de crecimiento exponencial en los dos sistemas, en el que se ha supuesto una tasa de crecimiento constante.

Ambos sistemas presentan una alta correlación, lo que da pie a pensar en una semejanza, característica que se ha podido corroborar con los tests de las medias y de homogeneidad de sus varianzas, contrastándose con los mismos que las dos muestras proceden de una misma población.

El estudio del solapamiento, a igual que lo hiciera Gluck (1990), refleja un alto porcentaje del mismo entre WoS y Scopus, si bien Scopus, con un 92% de cobertura sobre WoS, supera ligeramente a WoS, que alcanza solo el 74% sobre Scopus (datos coincidentes con el porcentaje de revistas comunes).

De igual forma se comportan los dos sistemas, al considerar las revistas; un núcleo de 35 en WoS y 42 en Scopus, respectivamente, la mayoría comunes, que cubren 1/3 del total de la producción total, formando el resto de revistas una gran cola, todas ellas de baja productividad.

La alta dispersión concuerda, también, con una fuerte concentración; con tan solo el 20% de las revistas se cubre el 90% de la literatura (Índice Gini = 0,8).

5. BIBLIOGRAFÍA

- ARCHAMBAULT, E.; CAMPBELL, D.; GINGRAS, Y. y LARIVIÈRE, V. Comparing bibliometric statistics obtained from the Web of Science and Scopus. *Journal of the American Society for Information Science and Technology*. 2009, vol. 60, n° 7, p. 1320-1326.
- BEARMAN, T.C. y KUNBERGER, W.A. *A study of coverage overlap among fourteen major science and technology abstracting and indexing services*. Philadelphia: National Federation of Abstracting and Indexing Services, 1977.
- BOST, W.E. Test on abstracts journals. *Journal of Documentation*, 1968, vol. 24, p. 61.
- BOURNE, C.P.; KASSON, M.S. y NORTH, J.B. *Overlapping coverage of bibliography of agricultura by fifteen other secondary sources*. U.S. Government Research and Development Report IGC-PA-6939), 1969.
- DE Solla Price, D.J. *Little Science, Big Science*. Columbia, 1963. (Traducción: Hacia una Ciencia de la Ciencia, Barcelona: Ariel, 1973. Traductor: J. M. López Piñero).
- FINGERMAN, S. Web of Science and Scopus: Current features and capabilities. *Issues in Science and Technology Librarianship*, 2006, 48 (Fall). [online]. Disponible en: <<http://istl.org/06-fall/electronic2.html>> [Consulta: 20 de septiembre de 2007].

- GLUCK, M. A review of journal coverage overlap with an extensión to the definition of overlap. *Journal of the American Society for Information Science*, 1990, vol. 41, nº 1, p. 43-60.
- GOODMAN, D. y DEIS, L. Web of Science (2004 version) and Scopus. *The Charleston Advisor*, 2005, vol. 6, nº 3, p. 5-21. <<http://charlestonco.com/comp.cfm?id=43>>.
- GOODMAN, D. y DEIS, L. Update on Scopus and Web of Science. *The Charleston Advisor*, 2007, vol. 7, nº 3, p. 15-18. <<http://charlestonco.com/comp.cfm?id=55>>.
- HOOD, W.W. y WILSON, C.S. Overlap in bibliographic databases. *Journal of the American Society for Information Science and Technology*, 2003, vol. 54, nº 12, p. 1091-1103.
- JACSO, P. As we may search – Comparison of major features of the *Web of Science*, *Scopus*, and *Google Acholar* citation-based and citation – enhanced databases. *Current Science*, 2005, vol. 89, nº 9-10, p.1537-1547.
- LAGUARDIA, C. E-Views and reviews: Scopus vs. Web of Science. *Library Journal*, 2005 [online] [15 de enero, 2005]. 3p. <<http://www.libraryjournal.com/index.asp?layout=articlePrint&article.cfm?ID=CA491154>>.
- LORENZ, M.O. Methods of measuring concentration and wealth. *Journal of the American Statistical Association*, 1905, vol. 9, p. 209-219.
- MARTYN, J. Tests on abstracts journals: coverage overlap and indexing. *Journal of Documentation*, 1967, vol. 23, p. 45-70.
- MARTYN, J. y SLATER, M. Tests on abstracts journals. *Journal of Documentation*, 1964, vol. 20, p. 212-235.
- PAO, M.L. Term and citation retrieval: A field study. *Information Processing & Management*, 1993, vol. 29, nº 1, p. 95-112.
- POYER, R.K. Journal article overlap among *Index Medicus*, *Science Citation Index*, *Biological Abstracts*, and *Chemical Abstracts*. *Bulletin of the Medical Library Association*, 1984, vol. 72, p. 353-357.
- PULGARÍN GERRERO, A. y ESCALONA FERNÁNDEZ, M.I. Medida del solapamiento en tres bases de datos con información sobre ingeniería. *Anales de Documentación*, 2007, vol. 10, p. 335-344.

ANEXO

Descripción del modelo matemático de crecimiento (Modelo exponencial)

La producción científica (número de publicaciones), en ingeniería química, en un determinado instante de tiempo, se indica como $P(t)$.

Entonces, $P(t + \Delta t)$ sería la producción científica en el instante $t + \Delta t$.

Así, el crecimiento neto del número de artículos, en el intervalo de tiempo $[t, t + \Delta t]$, será igual a $P(t + \Delta t) - P(t)$ (1)

¿Podríamos calcular cuantas publicaciones han aparecido por unidad de tiempo, durante ese intervalo?

$$\frac{P(t + \Delta t) - P(t)}{\Delta t}$$

La respuesta sería la solución de la expresión $\frac{P(t + \Delta t) - P(t)}{\Delta t}$ (2)

Si al cociente anterior se le halla el límite, cuando el Δt se hace una cantidad muy pequeña, próxima a cero, se obtiene la derivada de la función $P'(t)$.

Con el concepto de derivada se puede explicar el desarrollo dinámico de la mayoría de los fenómenos naturales, pudiéndose extender a problemas de cualquier tipo de crecimiento.

$$\lim_{\Delta t \rightarrow 0} \frac{P(t + \Delta t) - P(t)}{\Delta t} = P'(t)$$

Se escribe de la siguiente forma: $\lim_{\Delta t \rightarrow 0} \frac{P(t + \Delta t) - P(t)}{\Delta t} = P'(t)$ (3)

Así que, la derivada, $P'(t)$, es el número de publicaciones aparecidas por unidad de tiempo.

Una vez calculada la expresión (3), si se considera un intervalo de tiempo muy

pequeño ($\Delta t \rightarrow 0$), esta se convierte en $\frac{P(t + \Delta t) - P(t)}{\Delta t} \cong P'(t)$ (4)

La expresión (4) se puede transformar en $P(t + \Delta t) - P(t) \cong P'(t) \cdot \Delta t$.

Y, finalmente, en $P(t + \Delta t) = P(t) + P'(t) \cdot \Delta t$. (5)

La expresión (5) se interpreta como “el número de artículos existentes en el tiempo $t + \Delta t$ es igual a los que había en el tiempo t más los que han sido publicados en el intervalo $[t, t + \Delta t]$ ”. Los que aparecen en el intervalo $[t, t + \Delta t]$, es igual al producto $P'(t) \cdot \Delta t$, esto es, los que aparecen por unidad de tiempo multiplicados por el tiempo.

Por otro lado, la *tasa de crecimiento media relativa*, será la variación neta del número de artículos, por unidad de tiempo (por ejemplo un año), dividida por el número de artículos que teníamos a comienzo de ese año.

$$\frac{\frac{P(t + \Delta t) - P(t)}{\Delta t}}{P(t)} = \frac{P(t + \Delta t) - P(t)}{P(t) \cdot \Delta t}$$

Y la *tasa de crecimiento instantánea*, que mide la tendencia del crecimiento, de la producción científica bajo estudio, en el instante t , se obtiene calculando el límite de la expresión anterior, cuando $\Delta t \rightarrow 0$.

$$\frac{1}{P(t)} \cdot \lim_{\Delta t \rightarrow 0} \frac{P(t + \Delta t) - P(t)}{\Delta t} = \frac{1}{P(t)} \cdot P'(t) = \frac{P'(t)}{P(t)} = a$$

$$P'(t) = a \cdot P(t)$$
 (6)

La expresión (6) es una ecuación diferencial ordinaria, que puede resultar algo complicada de resolver, en los casos donde la tasa de crecimiento, a , no sea constante.

La hipótesis más simple es cuando la tasa de crecimiento es constante (modelo lineal).

Esto significa que cualquiera que sea el tiempo, para un intervalo de tiempo pequeño, se tiene $P(t + \Delta t) - P(t) = a \cdot P(t) \cdot \Delta t$. (7)

Esta expresión (7) no depende del tamaño de la producción científica. Se trata de de una ecuación lineal de fácil estudio.

En primer lugar, se puede comprobar que todas las funciones de la forma $P(t) = P(0) \cdot e^{at}$, con $P(0)$ como constante arbitraria, son soluciones de (6).

$$\frac{d}{dt} [P(0) \cdot e^{at}] = P(0) \cdot a \cdot e^{at} = a \cdot P(t), \text{ para todo } t \in R.$$

Así que, si tenemos una producción científica inicial, de $P(0) > 0$, suponiendo que $a > 0$, crecerá de acuerdo con la fórmula $P(t) = P(0) \cdot e^{at}$. (8)

Se trata de un crecimiento ilimitado (maltusiano), cuyo $\lim_{t \rightarrow \infty} P(t) = \infty$.

$$\frac{P'(t)}{P(t)} = a = \frac{d \ln P(t)}{dt}$$

También podemos partir de (6):

Para llegar a la expresión anterior hay que tener en cuenta que

$$\ln P(t) = a \cdot t + c \tag{9}$$

Aplicando la definición de \ln en (9), se obtiene: $P(t) = e^{a \cdot t + c} = e^{a \cdot t} \cdot e^c = C \cdot e^{a \cdot t}$

Pero la constante C coincide con $P(0)$, así que, finalmente nos resulta que la ecuación del crecimiento exponencial es $P(t) = P(0) \cdot e^{at}$. (8)

El cálculo del número de artículos que aparecen por unidad de tiempo (año) se obtiene de la siguiente forma:

El número de artículos en el tiempo $t + \Delta t$ será $P(t + \Delta t) = P(0) \cdot e^{a(t + \Delta t)}$.

Y el crecimiento efectivo durante ese tiempo será:

$$P(t + \Delta t) - P(t) = P(0) \cdot (e^{a(t + \Delta t)} - e^{a \cdot t})$$

Por lo tanto, el número de artículos que aparecen por año en el intervalo $[t, t + \Delta t]$ se obtiene como sigue:

$$\frac{P(t + \Delta t) - P(t)}{\Delta t} = P(0) \frac{(e^{a(t + \Delta t)} - e^{a \cdot t})}{\Delta t} \tag{10}$$

Calculando el $\lim_{\Delta t \rightarrow 0} \frac{(e^{a(t + \Delta t)} - e^{a \cdot t})}{\Delta t}$ (Regla de l'Hopital):

$$\begin{aligned} \lim_{\Delta t \rightarrow 0} \frac{(e^{a(t + \Delta t)} - e^{a \cdot t})}{\Delta t} &= \lim_{\Delta t \rightarrow 0} \frac{e^{a \cdot t} (e^{a \cdot \Delta t} - 1)}{\Delta t} \text{ (l'Hospital)} = \lim_{\Delta t \rightarrow 0} \frac{e^{a \cdot t} (a \cdot e^{a \cdot \Delta t})}{1} = \\ &= e^{a \cdot t} \cdot \lim_{\Delta t \rightarrow 0} a \cdot e^{a \cdot \Delta t} = a \cdot e^{a \cdot t} \end{aligned} \tag{11}$$

Multiplicando el resultado del límite por $P(0)$, esto es, $P(0) \cdot a \cdot e^{a \cdot t}$, se obtiene la derivada de $P(t) = P(0) \cdot e^{a \cdot t}$ (8), que es igual al número de artículos por unidad de tiempo.

En cuanto al tiempo de duplicación de la literatura científica en ingeniería química, se trata de calcular el tiempo que tarda en multiplicarse por 2 la cantidad con la que partimos, esto es, $P(t)$ en $2 \cdot P(t)$.

$$\text{Por tanto } \frac{2P(t)}{P(t)} = e^{a \cdot t}; 2 = e^{a \cdot t}; \ln 2 = a \cdot t \cdot \ln e; 0,6931 = a \cdot t; t = \frac{0,6931}{a}.$$

Si indicamos con T el tiempo de duplicación, la expresión para calcularlo sería

$$T = \frac{\ln 2}{a} \quad (12).$$