

MEDIDA DEL SOLAPAMIENTO EN TRES BASES DE DATOS CON INFORMACIÓN SOBRE INGENIERÍA

*Antonio Pulgarín Guerrero**

Departamento de Informática. Universidad de Extremadura.

*M^a. Isabel Escalona Fernández***

Documentación. Biblioteca Central. Universidad de Extremadura.

Resumen: El contenido de las bases de datos suele variar tanto en diversidad temática como en extensión; unas cubren información específica, mientras que otras son multidisciplinarias. Por eso es difícil seleccionar cual o cuales se mantienen o eliminan del repertorio de un servicio de información, cuando entran en juego factores económicos o de otra índole. El estudio del solapamiento y/o la cobertura de una serie de bases de datos corresponden a la Informetría. Los resultados serán fundamentales a la hora de tomar una decisión al respecto. El presente trabajo lleva a cabo un estudio del solapamiento y/o cobertura que presentan tres bases de datos con contenidos en ingeniería (IEL, INSPEC y PASCAL). Se emplearon índices de solapamiento y de semejanza (Jaccard y de Meyer entre otros) para comparar estas tres bases de datos. Los resultados muestran una gran diferencia entre las bases de datos, tanto en cobertura como en solapamiento. INSPEC es la base de datos con mayor cobertura (aproximadamente el 50 % de los documentos y más del 50 % de las fuentes) y un solapamiento con las otras dos bases de datos superior al 30 %.

Palabras clave: Medida; bases de datos; IEL; INSPEC; PASCAL; solapamiento; ingeniería; informetría.

Title: MEASUREMENT OF THE OVERLAPPING IN THREE DATA BASES WITH INFORMATION ON ENGINEERING.

Abstract: The content of the databases usually varies so much in thematic diversity as in extension; some data bases cover specific information, whereas others are multidisciplinary. For that reason it is difficult to select as or as they stay or they eliminate of the collection of an information service, when economic factors take part or of another nature. The study of the overlapping and/or the cover of a series of data bases it corresponds to the field of informetrics. The results will be fundamental at the time of making a decision on the matter. The present work carries out a study of the overlapping and/or cover that present three data bases with contents in engineering (IEL, INSPEC and PASCAL). Similarity and overlapping indices were used (Jaccard and Meyer among others) to compare these three data bases. The results show a great difference between the data bases, as much in cover as in overlapping. INSPEC were the data base with greater cover (approximately 50% of documents and more of 50% of the sources) and an overlapping with the other data bases superior to 30%.

Keywords: Measurement; databases; IEL; INSPEC; PASCAL; overlapping; engineering; informetrics.

* pulgarín@unex.es

** escalona@unex.es

INTRODUCCIÓN

No es fácil decidir la eliminación de suscripciones a bases de datos del repertorio de un servicio de información, ya que la información contenida en una base de datos, más aún si es especializada, es fundamental para el usuario que la utiliza frecuentemente. Cuando una institución opta por aplicar un grado severo de exhaustividad pierde cobertura y extensión en información científica. Por otro lado hay que considerar el costo de las bases de datos, casi siempre alto, añadido al presupuesto general del servicio. Por ello se hace necesario estudiar la posibilidad de reducir los gastos sin disminuir la calidad de servicio al usuario.

Es aquí donde entra en juego la Informetría, con estudios de solapamiento de las bases de datos, como metodología, consistente en calcular el grado o la intensidad de semejanza que puede presentar un conjunto de estas bases.

Los estudios de solapamiento de bases de datos no son nuevos (Bearman y Kunberger, 1977), variando desde su descripción hasta la rigurosidad de su análisis (Gluck, 1990), y sobre temáticas muy variadas (Hood y Wilson, 2003; Parker, 2005).

Estos estudios se pueden agrupar bajo dos enfoques distintos; aquellos que se refieren sobre una definición teórica del solapamiento, presentando diversos índices con los que calcular dicho solapamiento; o los que se dedican a comprobar el grado de solapamiento entre bases de datos o fuentes secundarias.

Con este segundo enfoque se pretende observar la distribución de documentos o fuentes sobre un determinado tema por una serie de bases de datos (dispersión), o cómo es cubierta, esa misma literatura, por dos o más bases de datos (solapamiento). Esta información será muy útil a la hora de tomar decisiones sobre la adquisición, permanencia o eliminación de fuentes secundarias del repertorio de un servicio de información.

En el presente estudio se comparan tres bases de datos (IEL, INSPEC y PASCAL) con contenidos (se supone) similares, respecto al campo de la ingeniería. Se estudia el grado de solapamiento entre estas, con el fin de evitar gastos superfluos. Al mismo tiempo se podrá cuantificar la cantidad de información que se perdería en caso de eliminar una determinada base de datos.

Los cálculos sobre solapamiento se suelen hacer basándose en las fuentes primarias que cubren las fuentes secundarias o bien sobre los documentos (registros) que contenga sobre un tema determinado.

El primero de los procedimientos presenta una dificultad, consistente en la política de indexación de documentos que sigue cada base de datos; mientras unas vacían las fuentes completamente, otras lo hacen de forma selectiva (Pao, 1993). El segundo es mucho más laborioso para comparar las bases de datos. Además, los cálculos pueden hacerse sobre una población completa o sobre una muestra. En caso de utilizar muestras estas requieren ser seleccionadas con cuidado; una muestra seleccionada al azar debe proceder de una población uniforme (Provost, 1992).

MATERIAL Y MÉTODOS

El material de estudio para la realización de este trabajo se obtuvo de las bases de datos IEL, INSPEC y PASCAL, como las bases más idóneas y disponibles para el campo de investigación con el que se trabajó, la ingeniería.

IEEE/IEE Electronic Library (IEL), es una colección exhaustiva de documentos a texto completo que da acceso a publicaciones del IEEE (Institute of Electrical and Electronics Engineers) y del IEE (Institution of Electrical Engineers). Contiene más de un millón de artículos publicados en unas 12.000 publicaciones individuales (incluyendo artículos de revistas, normas y actas de congreso) especializadas en el campo de la ingeniería eléctrica, informática y electrónica. Su cobertura es desde 1988.

Information Services for Physics, Electronics and Computing (INSPEC), producida por The Institution of Electrical Engineering, es considerada como la principal fuente de información sobre física, ingeniería eléctrica y electrónica, informática y tecnologías de la información. Corresponde a las tres publicaciones impresas: *Physics Abstracts*, *Electrical and Electronics Abstracts* y *Computer and Control Abstracts*. Obtiene información de más de 4.250 revistas científicas y técnicas, de las que 750 se resumen por completo. También incluye actas de conferencias, libros, capítulos de libros, informes técnicos y tesis. Su cobertura es desde 1969.

PASCAL, es una de las principales fuentes mundiales de información científica y técnica, producida por el INIST (Institut de l'Information Scientifique et Technique). Corresponde a la publicación impresa *Bibliographie Internationale*. Incluye todo tipo de literatura: artículos de revistas (93%) del fichero, tesis (principalmente francesas), actas de conferencias, informes técnicos, libros y patentes. Son indizadas más de 6.500 revistas, 4.500 de éstas con resumen completo. Su contenido está distribuido del siguiente modo: un 31 % de Medicina, un 22% de la Biología, un 17% Ingeniería, un 15% Física, un 7% sobre Astronomía, un 5% Química y un 3% Tecnología.

El propósito de este estudio fue examinar la distribución y/o solapamiento de fuentes sobre una serie de temas de ingeniería, que sean de interés general y, al mismo tiempo, relevantes en la actualidad.

El primer paso consistió en buscar una serie de temas de investigación y a partir de estos obtener términos de búsqueda más específicos y pertinentes con los que poder trabajar.

Para ello se accedió al Institut for Scientific Information (ISI) Essential Science Indicators¹, archivo dinámico elaborado por el ISI y que abarca la última década de la producción científica, proporcionando datos estadísticos y tendencias de la ciencia provenientes de más de 8.500 publicaciones incluidas en los índices ISI. Por tanto, puede ser considerada como una herramienta multidisciplinar que incluye 22 campos de investigación, entre los cuales figura el de "ingeniería".

La primera búsqueda se realizó en los frentes de investigación, sobre ingeniería.

Research Fronts (Frentes de investigación) son grupos de documentos muy citados, conocidos como "documentos centrales", en un tema especializado. Para identificar un frente de investigación, ISI efectúa un análisis de agrupamiento de citas conjuntas de los documentos muy citados durante los últimos 5 años.

De estos frentes se observaron los títulos de los trabajos clasificados por su temática en el campo de la ingeniería.

Los términos de búsqueda obtenidos de la observación de los trabajos correspondientes a los frentes de ingeniería, fueron introducidos, en cada base de datos, en el campo título con el fin de evitar cualquier discrepancia entre las bases de datos.

¹ <<http://portal.isiknowledge.com/portal.cgi?DestApp=ESI&Func=Frame>>.

Se seleccionaron los cinco términos siguientes sobre ingeniería, basados en su generalidad y en su alto nivel de interés actual como investigación: 1) Delay? and Neural-Networks; 2) Brain-Computer and (interface or communication); 3) ALGAN/GAN and (hem or high-electron-mobility); 4) Carbon-nanotube? and eletrode?; 5) Pem-Fuel-Cell?

La búsqueda se limitó a los años 2001-2005, con el fin de eliminar la actualización de las bases de datos como un factor de confusión.

El segundo paso consistió en determinar, en cada una de las tres bases de datos, cuantas de las fuentes primarias, recuperadas para cada término de búsqueda, estaban presentes en cada base de datos, como fuentes únicas, y cuantas duplicadas en cada grupo de dos bases.

Para cada ecuación de búsqueda y en cada base de datos se anotaba la primera aparición de una fuente primaria, ignorándose las siguientes ocurrencias de esa misma fuente. Las listas de fuentes por temas de búsqueda fueron comparadas entre cada grupo de dos bases de datos, indicándose con un 1 la presencia y con un 0 la ausencia de esa fuente.

El grado de solapamiento entre bases de datos y/o la cobertura de las mismas se puede calcular por distintos índices o métodos:

1.- Cuando lo que se busca es un criterio que evalúe la cobertura de una base de datos sobre un tema determinado, se suele utilizar el Índice de Meyer. Este índice, también conocido como índice relativo de peculiaridad, se utiliza en bases de datos mediante la repetición, en todas ellas, de un conjunto de búsquedas (Cañedo Andalia, 1999).

Imaginemos una base de datos, A, conteniendo 1.000 fuentes primarias, de las cuales 500 son fuentes únicas, 300 están duplicadas en otras bases de datos y 200 triplicadas. El cálculo del Índice de Meyer para esta base de datos sería:

$$IM_A = \frac{200 \times 0,3 + 300 \times 0,5 + 500 \times 1}{1.000} = \frac{60 + 150 + 500}{1.000} = 0,71$$

Con este criterio, las fuentes primarias únicas, contenidas en una base de datos, son las que tendrían un mayor peso, mientras que las duplicadas, por ejemplo, tendrían que dividirse por dos, las triplicadas por 3, etc. El resultado se puede interpretar como la cobertura que tiene una base de datos sobre una especialidad o temática determinada. Cuanto mayor sea el Índice Meyer mayor cobertura presenta la base de datos.

2.- A fin de evitar sesgos en muestras de tamaño pequeño, se suele utilizar el índice de Jaccard, usando la fórmula siguiente:

$$S_j = \frac{a}{(a + b + c)}, \text{ o bien } \frac{A \cap B}{(A \cup B) - (A \cap B)},$$

donde a es el número de fuentes comunes a dos bases de datos (A y B), b es el número de fuentes únicas en la primera base de datos y c es el número de fuentes únicas en la segunda base de datos.

El coeficiente resultante se puede multiplicar por 100 y expresarlo en porcentaje.

Un método parecido al de Jaccard es el solapamiento tradicional de cobertura entre dos fuentes secundarias, definido por Gluck (1990) y citado por Hood y Wilson (2003).

$$\% \text{ solapamiento} = 100 \times \left(\frac{|A \cap B|}{|A \cup B|} \right)$$

Al porcentaje más alto le corresponderá la mayor semejanza entre los resultados. Por ejemplo un coeficiente de 0,3 puede expresarse como el 30 % de semejanza o 70 % de diferencia. Estos datos significarían que, al buscar en una sola de las dos bases de datos, nos encontraríamos con un 70 % de fuentes únicas.

3.- Para medir el solapamiento o % de cobertura de una base de datos, A, respecto a otra, B, se emplea el % de solapamiento relativo, medida usada originalmente por Bearman y Kunberger (1977), definida posteriormente por Gluck (1990) y citados por Hood y Wilson (2003).

$$\% \text{ solapamiento en } A = 100 \times \left(\frac{|A \cap B|}{|A|} \right),$$

$$\% \text{ solapamiento en } B = 100 \times \left(\frac{|A \cap B|}{|B|} \right).$$

Con este método se tendría en cuenta el peso de las fuentes solapadas respecto a las de presencia única.

RESULTADOS

La consulta a las tres bases de datos arrojó un total de 2.197 documentos. Su distribución, por base de datos y por término de búsqueda, se puede observar en la tabla I. El término con mayor número de documentos en las tres bases de datos fue Delay? and Neural-Networks, con 787, y la base de datos en la que más documentos se recuperaron fue INSPEC, con 1.087, cerca del 50 % del total (49,48 %).

Base de datos	(1)*	(2)	(3)	(4)	(5)	Total	%
IEL	157	104	23	6	48	338	15,38
INSPEC	401	102	143	141	300	1,087	49,48
PASCAL	229	30	91	174	248	772	35,14
Total	787	236	257	321	596	2.197	100

* (1) Delay? and Neural-Networks

(2) Brain-Computer and (interface or communication)

(3) ALGAN/GAN and (hem or high-electron-mobility)

(4) Carbon-nanotube? and electrode?

(5) Pen-Fuel-Cell?

Tabla I. Distribución de documentos por base de datos y por términos de búsqueda.

Contabilizadas las fuentes que contenían los documentos, su distribución, por base de datos y por término de búsqueda, se muestran en la tabla II (columnas 2 a 5). Además, en esa misma tabla se pueden observar los resultados de la consulta, para cada una de las

base de datos, buscando todos los términos conjuntamente (columna 7), así como su porcentaje (columna 8). Los datos son similares a los de la tabla I (documentos), sobre todo, en los porcentajes; el mayor número de fuentes corresponde a la base de datos INSPEC (51,1 %).

Base de datos	(1)*	(2)	(3)	(4)	(5)	G**	%
IEL	68	29	17	6	35	142	20,79
INSPEC	131	44	39	65	106	349	51,10
PASCAL	58	21	21	48	62	193	28,11
Total						683	100

* Los números entre paréntesis del (1) al (5) se corresponden con los cinco términos de la búsqueda, tal y como se indica en la tabla I.

** G: número de fuentes primarias resultante de la consulta global ((1) a (5)), a cada una de las bases de datos.

Tabla II. Distribución del número de fuentes primarias resultante de la consulta a las tres bases de datos, por término de búsqueda y globalmente.

Los resultados del % de cobertura de una base de datos respecto a cada una de las otras dos, o solapamiento relativo, calculados a continuación, se muestran en la tabla III.

$$\% \text{ solapamiento en IEL} = \frac{IEL \cap INSPEC}{IEL} = \frac{108}{142} = 0,76$$

$$\% \text{ solapamiento en IEL} = \frac{IEL \cap PASCAL}{IEL} = \frac{28}{142} = 0,20$$

$$\% \text{ solapamiento en INSPEC} = \frac{INSPEC \cap IEL}{INSPEC} = \frac{108}{349} = 0,31$$

$$\% \text{ solapamiento en INSPEC} = \frac{INSPEC \cap PASCAL}{INSPEC} = \frac{131}{349} = 0,38$$

$$\% \text{ solapamiento en PASCAL} = \frac{PASCAL \cap IEL}{PASCAL} = \frac{28}{192} = 0,15$$

$$\% \text{ solapamiento en PASCAL} = \frac{PASCAL \cap INSPEC}{PASCAL} = \frac{131}{192} = 0,68$$

Bases de datos	IEL	INSPEC	PASCAL
IEL	-----	0,76 (108)	0,20 (28)
INSPEC	0,31 (108)	-----	0,38 (131)
PASCAL	0,68 (131)	0,15 (28)	-----

Tabla III. Solapamiento relativo o % de cobertura de cada base de datos respecto a cada una de las otras.

Observando la primera columna de la tabla III se aprecia que IEL cubre, con 108 fuentes en común, tan solo el 31 % del contenido de INSPEC, mientras que si observamos la columna segunda de la misma tabla se aprecia que INSPEC, con el mismo número de fuentes en común, cubre el 76 % del contenido de IEL.

Por lo tanto, es la base de datos INSPEC la que obtiene una mayor cobertura de fuentes sobre ingeniería. PASCAL (columna 3) cubre el 20 % de IEL y el 38 % de INSPEC.

El grado de semejanza entre las bases de datos, agrupadas de dos en dos, fue calculado aplicando el índice de Jaccard o índice de semejanza a los distintos grupos de bases de datos y para cada término de búsqueda, así como su media. Los resultados se muestran en la tabla IV.

Base de datos	(1)*	(2)	(3)	(4)	(5)	Media
IEL - INSPEC	0,28	0,49	0,24	0,09	0,27	0,28
IEL - PASCAL	0,09	0,09	0,19	0,02	0,10	0,10
INSPEC - PASCAL	0,32	0,16	0,46	0,28	0,37	0,32
Media	0,23	0,25	0,30	0,13	0,25	

Tabla IV. Índice de Jaccard o de semejanza entre los distintos grupos de bases de datos, distribuidos por término de búsqueda y media resultante.

Como se puede observar existen diferencias a tener en cuenta entre los distintos índices.

Los coeficientes resultantes se distribuyen entre 0,02 (Grupo de bases de datos IEL – PASCAL, para el término de búsqueda Carbon-nanotube? and electrode?) y 0,49 (Grupo de bases de datos IEL – INSPEC, para el término de búsqueda Brain-Computer and interface or communication), lo que indica que ningún grupo compartió menos del 2 % de las fuentes ni más del 50 %.

Examinando los resultados con más detalles observamos que el término ALGAN/GAN and (hem or high-electron-mobility) obtiene la mayor semejanza (media = 0,30) para todas las parejas de bases de datos, mientras que el término Pen-Fuel-Cell? fue el menos semejante (0,13). Por pareja de bases de datos, considerando todos los términos de búsqueda globalmente, INSPEC – PASCAL obtiene la mayor media (0,32) seguida de IEL – INSPEC (0,28) y la que menos semejanza tuvo fue la pareja formada por IEL – PASCAL con solo el 10 % de semejanza.

Con objeto de evaluar la exhaustividad de la cobertura de las bases de datos sobre la temática propuesta, teniendo en cuenta el número de fuentes únicas (las de mayor peso), el de fuentes duplicadas con cualquiera de las otras dos y el de triplicadas, se procedió al cálculo del Índice de Meyer. Los resultados se muestran a continuación.

$$\text{Base de Datos IEL} = \frac{28 \times 0.3 + 80 \times 0.5 + 34 \times 1}{142} = 0.58$$

$$\text{Base de Datos INSPEC} = \frac{28 \times 0.3 + (80 + 113) \times 0.5 + 156 \times 1}{349} = 0.75$$

$$\text{Base de Datos Pascal} = \frac{28 \times 0.3 + 113 \times 0.5 + 51 \times 1}{192} = 0.6$$

Los resultados indican que la base de datos con mayor exhaustividad fue INSPEC, con un 75 % de cobertura, seguido de PASCAL, con un 60 %, y de IEL, con un 58 %.

Finalmente, se muestran los resultados de la búsqueda respecto al número de fuentes solapadas entre los distintos grupos (tabla V).

Base de datos	(1)	(2)	(3)	(4)	(5)
IEL - INSPEC	44	24	11	6	30
IEL - PASCAL	10	4	6	1	9
INSPEC - PASCAL	46	9	19	25	45

Tabla V. Distribución del número de fuentes primarias duplicadas, por pareja de bases de datos y por término de búsqueda.

En la tabla se aprecia que el grupo INSPEC – PASCAL es el que tiene mayor número de fuentes en común, seguido por IEL – INSPEC, siendo muy reducido el número de fuentes compartidas entre las bases de datos IEL y PASCAL.

DISCUSIÓN Y CONCLUSIONES

La idea que al principio se planteaba era cuantificar el solapamiento y cobertura de la literatura contenida en una serie de bases de datos con objeto de tener criterios que ayudasen a la toma de decisiones respecto a la adquisición, mantenimiento o eliminación de bases de datos del repertorio de un servicio de información.

Poder comparar una serie de bases de datos, con metodología procedente de la Informática, a fin de evitar errores no deseados a la hora de discernir cuál o cuáles bases de datos cubriría una mayor cantidad de literatura científica era el objetivo planteado. Con los métodos utilizados, los resultados nos ayudarían a una mejor comprensión del problema.

No es fácil comparar bases de datos como las estudiadas en este trabajo, cada una con un volumen de literatura distinto y muy diverso, aunque las tres con contenido abundante en el campo de la “ingeniería”.

Con el cálculo de los índices utilizados se obtuvo una aproximación del grado de solapamiento entre las bases de datos, de la semejanza entre ellas y, también, del porcentaje de cobertura que alcanzaron cada una por separado. Con estos criterios se podría tener una visión global del problema y ayudaría a tomar una decisión correcta; por un lado evitando dejar al usuario sin una información fundamental para su investigación y por otro reduciendo gastos superfluos.

Se ha elegido, como material de estudio, las fuentes primarias contenidas en las bases de datos, si bien se podría haber utilizado los documentos, pero estos últimos requieren un trabajo mucho más laborioso y los resultados no serían muy diferentes.

La primera conclusión que se obtiene es que este aporta una metodología que se puede utilizar para casos similares de solapamiento de bases de datos.

Otra segunda conclusión es la comprobación de que existe una gran cantidad de fuentes únicas en las bases de datos estudiadas, como se puede apreciar en el solapamiento relativo, a pesar de tener un alto contenido en “ingeniería” las tres. Este hecho indica que

se podría perder un significativo número de registros si se consultase cualquier otra base de datos.

La tercera conclusión no es realmente una conclusión, sino una hipótesis. Ciertos autores han especulado con el significado de que haya un alto número de registros, en este caso fuentes, duplicados. Una hipótesis razonable sería aquella que dice que un alto porcentaje de registros duplicados es mucho más importante en cierto sentido. Se necesitaría profundizar más en este aspecto para comprobar si esto es cierto; por ejemplo, si los artículos o las fuentes altamente duplicados son también muy citados (Pao, 1993; Ingwersen y Christensen, 1997).

La cuarta conclusión es que del presente estudio se deduce que la base de datos INSPEC es la que más cobertura presenta, indicando con ello que para el campo de la "ingeniería" es fundamental. Este es el motivo por el que presenta poca semejanza con las otras dos bases de datos y por el que obtiene un alto solapamiento relativo (76 % de IEL y 68 % de PASCAL).

Por último, a modo de conclusión final, se podría decir que en el caso de prescindir de alguna de las tres bases de datos, debería ser IEL. Esta base de datos, aunque contiene los documentos a texto completo, debería tratarse con cautela tal característica, ya que hoy día, existen mecanismos suficientes, proporcionados por las tecnologías de la información que hacen posible la recuperación del texto completo de cualquier documento. No se puede sacrificar una amplia cobertura temática o el acceso a una gran cantidad de información por un número mucho más reducido de documentos a texto completo. El número final de usuarios que potencialmente podría usar una y otra base de datos sería muy diferente.

REFERENCIAS

- 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.
- CAÑEDO ANDALIA, R. Estudio de solapamiento en la selección de las publicaciones seriadas y las bases de datos. *ACIMED*, 1999, vol. 7, nº 3, p. 164-170.
- GLUCK, M. A review of journal coverage overlap with a extension to the definition of overlap. *Journal of the American Society for Information Science*, 1990, vol. 41, nº 1, p. 43-60.
- 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.
- INGWERSEN, P. y CHRISTENSEN, F. H. Data set isolation for bibliometric online analyses of research publications: Fundamental methodological issues. *Journal of the American Society for Information Science*, 1997, vol. 48, p. 205-217.
- ISI Web of Knowledge, Essential Science Indicators. (2006) [en línea]. Disponible en: <<http://portal.isiknowledge.com/portal.cgi?DestApp=ESI&Func=Frame>>. [Consultado: 15 octubre 2006].
- PAO, M. L. Term and citation retrieval: A field study. *Information Processing & Management*, 1993, vol. 29, nº 1, p. 95-112.
- PARKER, J. (2005, spring) [en línea]. *Evaluating bibliographic database overlap for marine science literature using an ecological concept: Issues in Science and Technol-*

ogy Librarianship. Disponible en: <www.istl.org/05-spring/refereed-2.html>. [Consultado: 12 septiembre 2006].

PROVOST, F. y NIEUWENHUYSEN, P. Measuring overlap of data bases in water supply and sanitation using sampling and the binomial probability distribution. *Scintometrics*, 1992, vol. 25, n^o 1, p. 201-209.