

COMPARACIÓN Y EVALUACIÓN DE LAS BASES DE DATOS ERIC, LISA E ISA SOBRE EL TEMA «RECUPERACIÓN DE LA INFORMACIÓN»

J. A. Salvador Oliván¹, J. M. Angós Ullate², M.ª J. Fernández Ruiz³

Resumen: Se ha realizado una búsqueda bibliográfica sobre «recuperación de la información» en las bases de datos ERIC, LISA e ISA del distribuidor DIALOG, durante el período de tiempo de enero de 1995-junio de 1997, exigiendo que apareciera el término sólo en el título. Los objetivos son estudiar las características y factores relacionados con la cobertura, actualización e indización de las bases de datos, así como determinar los temas de investigación más frecuentes y el coste-beneficio de las bases de datos. En los resultados de la evaluación, una característica común y negativa de las tres bases de datos es el considerable tiempo de demora en su actualización. Respecto a la indización, es LISA la que adolece de especificidad y exhaustividad en los términos índice, si bien es la que mayor número de revistas y de documentos cubre. Debido a la existencia de referencias no pertinentes, puede resultar la base de datos más cara. ERIC destaca por ser económica, buen grado de especificidad, exhaustividad y precisión en la indización; sin embargo cubre pocas revistas aunque son las más importantes. ISA es una base de datos con una cobertura importante e indizada con precisión y exhaustividad; por contra, es la más cara de todas.

Palabras clave: recuperación de la información; evaluación; bases de datos online; LISA; ERIC; ISA.

Summary: A bibliographical search in the databases ERIC, LISA and ISA of the DIALOG host, during the period January 1995 to June 1997, requiring that the term «information retrieval» appeared in the title has been carried out. The objectives are to study the characteristics and factors related to the coverage, update and indexation of the databases, as well as to determine the most frequent research topics and the cost-benefit of the databases. In the results of the evaluation, a common and negative characteristic of the three databases is the considerable time delay in their update. In relation to the indexation, LISA lacks specificity and exhaustivity in the index terms, although it is the one that covers a larger number of journals and documents. Due to the existence of non pertinent references, it can turn out to be the most expensive database. ERIC stands as being economic, with a high degree of specificity, exhaustivity and precision in the indexation; however it covers few journals although they are the most important. ISA is an accurate database with an important coverage and exact and exhaustive indexation, but it is the most expensive of them all.

Keywords: information retrieval, evaluation, online databases, LISA, ERIC, ISA.

¹ Departamento de Ciencias de la Documentación e Historia de la Ciencia. Facultad de Filosofía y Letras. Universidad de Zaragoza. Correo electrónico: jaso@posta.unizar.es.

² Centro de Documentación Científica. Universidad de Zaragoza. Correo electrónico: angos@posta.unizar.es.

³ Ayuntamiento de Zaragoza. Correo electrónico: mjferuiz@posta.unizar.es.

Recibido: 12-6-98.

1 Introducción

Para el desempeño de cualquier actividad profesional (educativa, investigadora, laboral, etc.) se hace necesario un estado del conocimiento sobre el tema en cuestión en constante actualización, lo que puede representar un quebradero de cabeza teniendo en cuenta el exceso de producción científica y de fuentes de información que pueden existir. De todos es conocido que la información científica se organiza hoy en día en forma de bases de datos, bien online o en CD-ROM, permitiendo una búsqueda exhaustiva y actual sobre cualquier tema. La selección de aquellas que mejor se ajusten a la necesidad de información del usuario será un factor clave para satisfacer con éxito la demanda de información.

Por otro lado, hay que señalar que para un tema específico normalmente suelen existir varias bases de datos, lo que puede plantear el problema de su selección. A veces, y en función del motivo que lleve al usuario a solicitar una búsqueda bibliográfica, será necesario consultar diferentes bases de datos para cubrir con exhaustividad el tema; sin embargo, en otras ocasiones, quizá sea suficiente la consulta de sólo una base de datos. En el proceso de selección será más fácil tomar una decisión acertada si conocemos determinadas características de las bases de datos, como cobertura, tipo de documentos que contienen, actualización, existencia o no de vocabulario controlado o tesoro, política de indización, etc.; estas características definirán el grado de calidad de las bases de datos, y por consiguiente, en alguna medida, la calidad de los resultados de la búsqueda sobre un determinado tema (1, 2).

En el campo de la Documentación, existen varios trabajos donde se estudian algunas de estas características en las bases de datos propias de nuestra especialidad, fundamentalmente orientados a evaluar y comparar las prácticas de indización en las bases de datos online LISA, ISA y LL (3), ERIC Y LISA (4, 5, 6), ERIC, LISA e ISA (7, 8) o LISA en formato CD-ROM (9).

2 Objetivos

Una de las principales áreas de investigación que actualmente existen en nuestro campo es la relacionada con la recuperación de información, lo que ha provocado en los últimos años un aumento importante, tanto en el número de publicaciones como en la aparición de revistas especializadas sobre el tema. El propósito general de este trabajo es conocer las principales líneas de investigación en esta área a partir de los descriptores utilizados con mayor frecuencia en la indización de los artículos y examinar las características de las bases de datos online más representativas sobre el tema «recuperación de la información» en el ámbito de las Ciencias de la Documentación, para:

- evaluar la calidad de dichas bases de datos, atendiendo a los factores de cobertura, actualización, indización, y
- proporcionar un conocimiento al documentalista que le permita elegir la base de datos más adecuada y económica, tanto desde el punto de vista de satisfacer sus propias necesidades de información como desde el punto de vista de un usuario en particular.

Las bases de datos elegidas para llevar a cabo estos objetivos han sido las siguientes: LISA (Library and Information Science Abstracts), ISA (Information Science Abstracts) y ERIC (Educational Resources Information Center). Las razones que nos han llevado a seleccionar estas bases de datos han sido las siguientes: LISA e ISA son las más representativas e importantes en el ámbito de la ciencia de la información y documentación; la elección de ERIC viene determinada por constituir la principal fuente de información en educación y ser de utilidad tanto para profesores como estudiantes de Biblioteconomía y Documentación, por indizar las revistas de mayor impacto en Ciencias de la Documentación y por ser una base de datos mucho más barata que LISA o ISA, lo que puede repercutir en un coste económico menor para el usuario cuando ante una búsqueda exhaustiva se utilizan estas tres bases de datos de forma conjunta; por último, otra razón muy importante es que ERIC es de acceso gratuito a través de la Web, de manera que al profesional de la Documentación le puede resultar muy útil consultar esta fuente de manera aislada para conseguir información de interés en su especialidad.

3 Metodología

- El primer paso ha consistido en plantear una búsqueda de información sobre «recuperación de la información» en el distribuidor norteamericano DIALOG en las bases de datos LISA, ISA y ERIC.
- En segundo lugar, se interrogó en las citadas bases de datos por la frase INFORMATION (W) RETRIEVAL, búsqueda que se realizó en julio de 1997. Se limitó a que aparecieran dichos términos en el título para conseguir una gran precisión, de manera que presumiblemente aquellos registros recuperados serían idóneos y altamente pertinentes para poder averiguar en qué medida el concepto de recuperación de información aparece también como descriptor y estudiar los fallos en la precisión y especificidad del vocabulario utilizado en cada base de datos. No se buscó en otros campos (descriptores, resumen) por considerar que estaríamos introduciendo un sesgo importante al recuperar bastantes documentos cuyo tema principal no fuera «recuperación de información» y/o que lo trataran de forma superficial, con lo que no serían útiles para los propósitos de este estudio.

El número total de referencias fue de 4.157, por lo que se redujo el período de búsqueda a las publicaciones realizadas en los años 1995 hasta julio de 1997 e introducidas en las bases de datos, quedando un total de 257 referencias. La reducción a este período de tiempo se debió a considerar que los dos últimos años y medio son un período de tiempo suficiente para analizar y descubrir las cuestiones principales, investigaciones recientes y avances que se producen en el campo de la recuperación de la información, y muy importante también, que equivalía a un coste económico razonable.

- Se creó una base de datos en la que se introdujo información extraída de cada una de las referencias bibliográficas. Por una parte, se seleccionaron variables que podemos clasificar como identificativas o descriptivas de cada referencia, como nombre de la revista, fecha de publicación, tipo de documento, y base de

datos en la que aparecía. Y por otra parte, se seleccionaron variables relacionadas con la parte temática y contenido del documento: descriptores que aparecen, presencia de *information retrieval* como descriptor; en este aspecto, también se evaluó si el título del artículo estaba completamente reflejado en los descriptores. Todas estas variables son necesarias por una parte para medir la calidad de las bases de datos, tanto en lo que se refiere a su cobertura, actualización y factores relativos a la indización, y por otra para representar los temas más investigados y analizar el coste-beneficio.

- Por último, se realizó el proceso de datos con el paquete estadístico SPSS al objeto de analizar los datos.

4 Resultados

Las características estudiadas de las bases de datos hacen referencia a diferentes aspectos cualitativos y factores que pueden afectar al rendimiento de éstas y a la calidad de los resultados finales de una búsqueda.

4.1 Factores de cobertura

Números de ítems: El número total de referencias obtenidas en las que aparecía *information retrieval* en el título ha sido de 257, siendo la base de datos LISA la que presenta mayor número de publicaciones (145), seguida de ISA (80) y por último de ERIC (32). Para el estudio de algunas características, se han desechado 17 ítems, todos ellos de la base de datos LISA, ya que son revisiones de libros, no son resultado de una actividad científica e investigadora y por consiguiente carecen de todo interés, y además algunos de ellos son repetitivos, llegando a aparecer el mismo resumen 5 veces en revistas diferentes. Queda, pues, finalmente un número total de 240 referencias, de las que 128 están en LISA, 80 en ISA y 32 en ERIC.

Unicidad y grado de solapamiento: Existe un grado de solapamiento global del 34,6%. Esto quiere decir que eliminando los ítems duplicados en las bases de datos consultadas, nos quedan 157 referencias de documentos únicos introducidos en las bases de datos durante 2 años y medio.

De las 157 referencias, 26 aparecen en las tres bases de datos, lo que representa un solapamiento entre las tres bases de datos de un 16,5%; entre ERIC e ISA, coinciden estas 26, y entre ERIC y LISA, estas 26 más otras 2 (28 en total); y por último, entre LISA e ISA, las 26 más 29 referencias más, lo que suma un total de 55 referencias (Tabla I). El mayor grado de solapamiento corresponde a las bases de datos LISA e ISA, algo lógico ya que cubren la misma área de conocimiento.

El número de referencias bibliográficas únicas y que no se repiten entre las bases de datos es de 100, lo que representa el 63,7%; ERIC sólo aporta 4 referencias únicas, mientras que ISA aporta 25 y LISA con 71 es la que mayor número de referencias únicas ofrece (Tabla II).

De forma global, ERIC con 32 referencias cubre el 20,4 % del total, ISA con 80 referencias, el 50,9 % y LISA con 128 referencias el 81,5 %. Desde otro punto de vista, podemos decir que el 87,5 % de las referencias que aparecen en ERIC, aparecen en

Tabla I
Solapamiento entre las bases de datos

	N.º	%	Total ítems
ERIC, LISA e ISA	26	16,5	157
ERIC y LISA	28	21,2	132
ERIC e ISA	26	30,2	86
ISA y LISA	55	35,9	153

Tabla II
Referencias únicas en las bases de datos

	N.º	%
ERIC	4	2,5
ISA	25	16,0
LISA	71	45,2
TOTAL	100	63,7

cualquiera de las otras dos bases de datos; el 68,7 % de las referencias de ISA, aparecen en ERIC y/o LISA; y por último, este porcentaje se reduce a un 44,5 % de referencias en LISA que aparecen también en ERIC y/o ISA. Esto nos confirma que es LISA la que contribuye con un mayor número de referencias que no aparecen en las otras bases de datos.

Tipo de documento: El tipo de documento que más frecuentemente aparece es el artículo de revista con un 75,2 %, seguido de las comunicaciones a congresos con un 11,5 % y las patentes con un 8,9 % (Tabla III). En la distribución por bases de datos, lo más destacable es que la base de datos ISA es la única que indiza patentes, y en ERIC aparece una tesis y un paper practicum.

Fuentes de publicación: La fuente en la que aparecen mayor número de referencias es la revista *Journal of the American Society for Information Science (JASIS)* con 17 artículos (10,8 %), seguida de *Information Processing & Management* con 13

Tabla III
Tipos de documentos

	ERIC		LISA		ISA		TOTAL	
	N.º	%	N.º	%	N.º	%	N.º	%
Artículos	25	78,1	106	82,8	54	67,5	118	75,2
Comunicaciones a congresos	3	9,4	17	13,3	10	12,5	18	11,5
Patentes	0	0,0	0	0,0	14	17,5	14	8,9
Revisiones	2	6,3	4	3,1	2	2,5	4	2,5
Tesis	1	3,1	0	0,0	0	0,0	1	0,6
Paper practicum	1	3,1	0	0,0	0	0,0	1	0,6
Informe	0	0,0	1	0,8	0	0,0	1	0,6
TOTAL	32	100	128	100	80	100	157	100

(8,3 %), *Proceedings of the ASIS Annual Meeting* con 6 (3,8 %), *ACM Transactions on Information Systems* con 5 (3,2 %) y *Proceedings of the Research Colloquium of the British Computer Society Information Retrieval Specialist Group* con 5 (3,2%). Vemos, pues, cómo el 29,3 % de los registros se concentra en sólo 5 publicaciones, estando el resto muy disperso en 71 publicaciones seriadas y patentes.

Por bases de datos, destaca cómo en ERIC todas las referencias están en 7 revistas (más la tesis y el paper practicum), y de ellas, el 84,5 % están recogidas en 5 revistas, todas ellas con factor de impacto en el campo de la ciencia de la información y bibliotecas, y de ellas 4 son las que mayor factor de impacto tienen. En la base de datos LISA y en ISA destaca que también se recogen publicaciones de las principales revistas, si bien sólo alcanzan el 47,8 % y el 53,9 % respectivamente, estando el resto disperso por muy diversas revistas, sobre todo en LISA en la que se vacían 61 revistas diferentes y actas de congresos, mientras que en ISA esta cifra es de 29 (Tabla IV).

Tabla IV

Publicaciones más frecuentes e importantes en las bases de datos

	ERIC		LISA		ISA	
	N.º	%	N.º	%	N.º	%
<i>JASIS*</i>	8	25,0	17	13,3	14	17,5
<i>Information Processing & Management*</i>	11	34,3	13	10,2	13	16,3
<i>Proceedings of the ASIS Annual Meeting*</i>	3	9,4	6	4,7	6	7,5
<i>Journal of Documentation*</i>	3	9,4	4	3,1	4	5,0
<i>Annual Review of Information Science and Technology*</i>	2	6,3	2	1,6	2	2,5
<i>Review of Information Science</i>	2	6,3				
<i>Journal of Academic Librarianship*</i>	1	3,1	1	0,8		
<i>ACM Transactions on Information Systems</i>			5	3,9		
<i>Proceedings of the Research Colloquium of the British Computer Society Inf. Retr.</i>			5	3,9		
<i>Artificial Intelligence Review</i>			4	3,1		
<i>Canadian Journal Inf. and Library Science</i>			1	0,8	2	2,5
<i>Online CD-ROM Review*</i>			2	1,6	1	1,3
<i>Online*</i>			1	0,8	1	1,3
TOTAL	30	93,8	61	47,8	43	53,9

* Revistas con factor de impacto.

Destacable también el hecho de que en LISA y en ISA, en las principales revistas prácticamente hay el mismo número de ítems, lo que significa que el tiempo que tardan en vaciar e indizar las revistas estas dos bases de datos es muy similar. Sin embargo, sorprende que para la revista más prestigiosa o de mayor impacto, como es JASIS, hay 3 ítems menos en ISA. Revisadas las fechas de publicación, no parece que la causa fundamental de esta discrepancia sea únicamente el tiempo de retraso en la actualización de la base de datos, sino que también las políticas de criterios de calidad al seleccionar los artículos sean diferentes, o bien se puede atribuir a algún olvido de los indizadores e introductores de datos, ya que de los 3 que faltan en ISA, uno corresponde a agosto de 1995, otro a julio de 1996, y el último a febrero de 1997. Sí parece que el tiempo de demora en la indización, vaciado y actualización de la base de datos sea el motivo por el que en ERIC haya una diferencia de 9 ítems menos que en LISA y 6 menos que en

ISA, ya que, por ejemplo, de 1995 están todos los ítems, y algunos de 1996, pero ninguno de 1997, a la vez que también puede que exista algún criterio de calidad de los trabajos, ya que en 1996 no se correlacionan con el orden de publicación en la revista.

Con respecto a la revista *Information Processing & Management*, en LISA e ISA están los mismos ítems, y en ERIC faltan 2, uno que corresponde al año 1997, y otro al año 1995. Esto viene a apoyar la tesis de que efectivamente existe mucha más demora en ERIC que en las otras dos bases de datos, y además puede existir algún criterio de calidad o de conveniencia para vaciar los artículos, algo totalmente comprensible tratándose de una base de datos de Educación, y en la que no es predominante el tema de Ciencias de la Información y Bibliotecas como en las otras dos bases de datos. Esto es corroborado también con los *Proceedings of the ASIS Annual Meeting*, en los que existen 3 publicados en 1995 y otros en 1996, pero en ERIC sólo están recogidos los del año 1995.

4.2 Factores temporales: tiempo de demora y actualización

Como tal entendemos el tiempo que tarda en aparecer en la base de datos una referencia de un artículo desde su publicación en la revista impresa; dicho con otras palabras, se trata de comprobar la actualización de las bases de datos, característica primordial para su validez y utilidad, ya que una base de datos no actualizada deja de tener interés sobre todo para aquellos temas en los que el usuario requiera un conocimiento lo más actual y reciente posible.

El número total de publicaciones introducidas en las bases de datos durante 1995 fue de 72, similar a 1996 con 76; pero destaca el hecho de que sólo haya registradas 9 desde enero hasta junio, ambos inclusive, de 1997 (Tabla V). Analizando la distribución por bases de datos, observamos que tanto en LISA como en ISA el porcentaje es muy similar a lo largo del período de estudio; sin embargo en ERIC es mayor el porcentaje en 1995 (62,5 %) que en 1996 (37,5 %) y ninguna en 1997, lo que evidencia que en ERIC hay un retraso más que considerable en su actualización con respecto a este tema, superior a los 6 meses en las principales revistas.

Con respecto al tiempo de retraso en actualizar LISA e ISA es considerable, ya que según se muestra en la Tabla VI, la publicación más reciente en LISA es de abril en la revista *Online CD-ROM Review*, y las anteriores son todas de febrero y de enero, tanto en LISA como en ISA. De esta manera, el retraso en indizar las revistas más

Tabla V
Distribución de ítems según año de publicación y bases de datos

	1995		1996		1997	
	N.º	%	N.º	%	N.º	%
ERIC	20	62,5	12	37,5		
LISA	59	46,1	61	47,7	8	6,3
ISA	37	46,3	36	47,4	7	8,8
TOTAL	72	45,9	76	48,4	9	5,7

Tabla VI
Publicaciones del año 1997 en LISA e ISA

	LISA	ISA
<i>Online CDROM Review</i>	Abril	
<i>JASIS</i>	Febrero	
<i>Telematics and Informatics</i>	Febrero	Febrero
<i>Journal of Intelligent Information Systems</i>		Febrero
<i>Journal of Documentation</i>	Enero	Enero
<i>JASIS</i>	Enero	Enero
<i>JASIS</i>	Enero	Enero
<i>Information Processing and Management</i>	Enero	Enero
<i>New Library World</i>	No consta	No consta

importantes es de, como mínimo, 4 meses, ya que se ha comprobado que en la revista *JASIS* existen artículos publicados con el término *information retrieval* en el título en los números de marzo y mayo, y en la revista *Information Processing & Management* en los números de marzo y mayo; en las revistas de menor categoría el retraso es superior a los 6 meses.

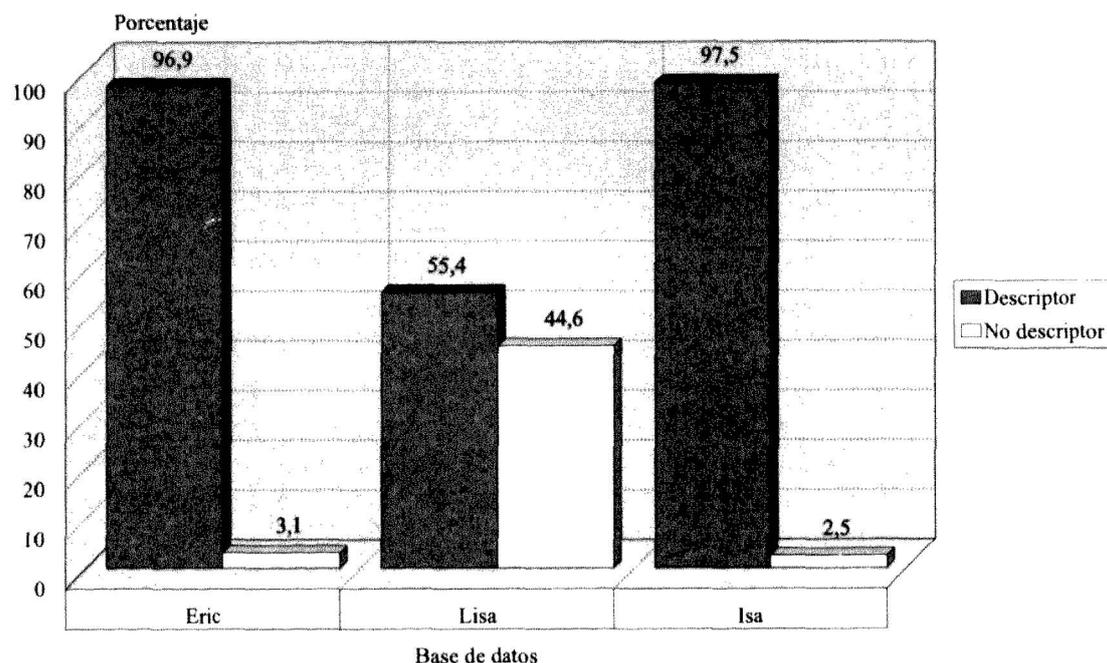
Así pues, nos encontramos con un problema importante como es el tiempo de demora, mucho más acentuado en ERIC que en LISA e ISA; sin embargo, en la primera podría considerarse como excusa que es una base de datos con prioridades en vaciado de revistas exclusivas de Educación; en las últimas es más difícil de excusar, sobre todo considerando que las revistas de mayor impacto en Ciencias de la Documentación sufren un retraso de más de 4 meses. A esto hay que añadirle el tiempo que tarda un artículo desde que se realiza el estudio hasta que se publica, aproximadamente 12 meses, lo que agrava todavía más el problema; con lo que nos encontramos que ante una demanda de información actual sobre un tema determinado, difícilmente pueden satisfacer estas bases de datos dicha exigencia.

4.3 Factores de indización y de vocabulario

Antes de analizar las cuestiones relacionadas con la indización, señalar que ERIC dispone de un tesoro además de contar con un campo de identificadores donde aparecen términos nuevos no contemplados en el tesoro; ISA utiliza un vocabulario controlado; y en LISA existe una lista de descriptores creada por un procedimiento de indización en cadena.

Especificidad en la asignación de términos: Es el factor más importante que influye en la precisión. La falta de especificidad puede causar fallos tanto en la exhaustividad como en la precisión de la búsqueda. La hemos medido a través de la presencia del término empleado en la búsqueda, y que pensamos que tiene entidad propia, en el campo de descriptores, lo que parecería la situación normal y deseable tanto si se usa un vocabulario controlado como si no. Esto no siempre se cumple, y así vemos en la figura 1 cómo en ERIC aparece en 31 ocasiones (96,9 %) y en ISA 78 (97,5 %), mientras que en LISA sólo figura como descriptor en 71 ítems (55,4 %), lo que representa poco más de la mitad. Existen algunos detalles dignos de mención:

Figura 1
Presencia de *Information Retrieval* como descriptor



- En ERIC se emplea el descriptor INFORMATION RETRIEVAL como término con entidad propia, y además aparece siempre como descriptor importante (marcado con un asterisco).
- En LISA aparece 3 veces como INFORMATION RETRIEVAL, 11 veces como COMPUTERIZED INFORMATION RETRIEVAL, 39 como ONLINE INFORMATION RETRIEVAL, 14 veces como COMPUTERIZED INFORMATION STORAGE AND RETRIEVAL, y 4 veces de otras maneras más vagas y extensas. Observamos cómo existe una gran variedad de términos para representar un concepto, además de que es la única base de datos en la que existe un descriptor en el que no aparecen las palabras INFORMATION y RETRIEVAL una a continuación de la otra, lo que debe tenerse en cuenta a la hora de utilizar operadores de proximidad.
- En ISA aparece como descriptor en forma de INFORMATION RETRIEVAL (57 registros) o de INFORMATION RETRIEVAL SYSTEMS (21 registros).

Un hecho a tener en cuenta es que si la búsqueda la hubiéramos realizado con la condición de que apareciera sólo en el campo de descriptores, hubiéramos perdido de forma individual en ERIC 1 ítem, en LISA 71 ítems y en ISA 2 ítems; en LISA se hubieran perdido 14 menos si se hubiera especificado el operador de sintaxis 2N, para así recuperar el término mencionado anteriormente.

Exhaustividad y Precisión en la indización: Se entiende por exhaustividad en la indización el grado en el se reconocen y representan todos los temas tratados en un documento. Por precisión la calidad de la indización por parte del indizador, que hemos medido observando si todos los conceptos del título están reflejados en los descripto-

res, y además son términos específicos (no deben confundirse con exhaustividad y precisión en la recuperación). Para ello, se ha analizado conceptualmente el título de cada artículo al objeto de determinar en qué grado se expresa el título en los descriptores. En la tabla VII se observa cómo en la base de datos ERIC en 27 ítems (84,4%), los descriptores reflejan todos los conceptos importantes del título, y de los 5 en los que sólo falta 1 concepto, en 4 de ellos no se recoge *hipertexto*, si bien se refleja como *hipermedia*. En la base de datos ISA, son 73 ítems (91,3 %) los que recogen todos los conceptos del título, por sólo 7 en los que falta un concepto. Sin embargo, en la base de datos LISA, solamente 38 registros (29,7 %) reflejan todos los conceptos importantes del título, mientras que el 61,7 % pierde algún concepto, y es la única base de datos en la que existen registros (11) con falta de 2 o más conceptos en los descriptores.

Tabla VII
Aparición de los conceptos del título
en los descriptores

	Total-mente		Falta un concepto		Falta más de un concepto	
	N.º	%	N.º	%	N.º	%
ERIC	27	84,4	5	15,6		
LISA	38	29,7	79	61,7	11	8,6
ISA	73	91,3	7	8,7		

Así pues, existe una diferencia importante entre las bases de datos ERIC e ISA, con un porcentaje importante de registros que contienen todos los conceptos importantes del título, por lo que podríamos considerar que están indizados exhaustivamente y con precisión, y la base de datos LISA, muy pobre en este aspecto, con un 70,3 % de registros en los que los descriptores reflejan de forma insuficiente el contenido de la publicación. Este aspecto consideramos que es muy importante, ya que si bien el título debe reflejar con precisión el contenido de la publicación y es responsabilidad del autor o autores, son los productores de las bases de datos quienes tienen la responsabilidad de indizar fielmente y exhaustivamente el tema de que tratan dichas publicaciones, y el título debe suponer el campo temático en el que deben poner más atención. De esto se puede deducir que en la base de datos LISA la exhaustividad de la indización es bastante pobre, y por consiguiente este factor disminuirá las tasas de precisión y exhaustividad en la recuperación.

Otros aspectos relevantes relacionados con la indización en LISA, y que contribuyen a tener una pobre impresión tanto por el vocabulario utilizado como por posibles errores de los indizadores, son los siguientes:

- Aparición de descriptores repetidos en una misma referencia: en varias referencias se observan secuencias de descriptores (en número de 4 o 5) que están duplicados.
- Aparece la abreviatura IR en el título, y en el campo de descriptores en lugar de aparecer como INFORMATION RETRIEVAL, el término índice es IR.
- Es curioso cómo en una referencia cuyo título es «.....information storage or in-

formation retrieval...», no está indizada con el descriptor COMPUTERIZED INFORMATION STORAGE AND RETRIEVAL, sino como COMPUTERIZED INFORMATION RETRIEVAL.

- También sorprende que en las revisiones de resúmenes de libros, los descriptores sean BOOK REVIEW ABSTRACTS y el título completo de la referencia.

Se ha calculado también el número de descriptores utilizados en las tres bases de datos, ya que es una medida relacionada con la exhaustividad o profundidad de la indización, y existe una relación directa entre el número de descriptores y la extensión con que los diferentes temas tratados en un documento se reconocen en la operación de indización. En la base de datos ERIC encontramos el valor más alto, con una media de 11,5 descriptores, seguida de ISA con 5,5 y LISA con 4,1 (Tabla VIII). Observamos, pues, cómo es en ERIC donde la profundidad de indización es más elevada, con un mínimo de 8 descriptores. Entre LISA e ISA hay una diferencia media de 1 descriptor más en la última, resultando un tanto extraña la presencia de registros en LISA en los que consta sólo 1 descriptor, además de que la media en esta base de datos está sesgada por la presencia de descriptores repetidos.

Tabla VIII
Media de descriptores utilizados en las bases de datos

	<i>Media ± D.E.</i>	<i>Mínimo</i>	<i>Máximo</i>
ERIC	11,5 ± 2,5	8	17
LISA	4,1 ± 1	1	12
ISA	5,5 ± 2	2	10

4.4 Análisis del contenido y temas de investigación

En cuanto a qué descriptores se utilizan con más frecuencia en las bases de datos, los podemos ver reflejados en la Tabla IX. Dejando aparte los relacionados con INFORMATION RETRIEVAL, parece claro que el descriptor más utilizado en las tres bases de datos es el que hace referencia al tema de búsquedas (SEARCHING, SEARCH STRATEGIES), que incluso en LISA es superior a recuperación de la información. Otros temas que coinciden entre algunas bases de datos son los que se refieren a indización, Internet, y evaluación.

4.5 Análisis coste-efectividad

A la hora de comparar coste-efectividad de la búsqueda realizada en las tres bases de datos, no se han contemplado todos los costes implicados en una búsqueda, y solamente se ha tenido en cuenta el coste por cita bibliográfica.

La base de datos más barata es ERIC (0,50\$/minuto de conexión y 0,65\$/cita bi-

Tabla IX
Descriptoros más utilizados en las bases de datos

<i>ERIC</i>	<i>LISA</i>	<i>ISA</i>
Information retrieval (31) Search strategies (11) Databases (7)	Searching (50) Online information retrieval (39) Computerized information storage and retrieval (14) Computerized information retrieval (11)	Information retrieval (57) Information retrieval systems (21) Searching (20)
Indexing (6)	Internet (11)	Databases (11)
Information processing (5) Knowledge representation (5) Relevance (5) Performance (5)	Evaluation (10) Software (9) Online databases (6)	Internet (8) Indexing (8) Evaluation (6) Artificial intelligence (5)

bibliográfica) seguida de LISA (0,50\$/minuto y 1,05\$ la cita) y la más cara ISA (1,25\$/minuto y 1,15\$/cita). Observamos en la Tabla X cómo el coste más alto por cita bibliográfica corresponde a ISA, si bien al desechar 17 citas de la base de datos LISA (revisiones de libros) y que no son pertinentes, va a causar que el coste por cita bibliográfica sea mayor en LISA.

Otro factor relacionado con el coste de la búsqueda es aquél que produce fallos en el comando RD (REMOVE DUPLICATE), debido a errores de mecanografía en el título o autores, y que repercutirá en la presencia de referencias que sólo deberían aparecer una sola vez y en un mayor coste. Así, hemos observado cómo el comando RD ha funcionado perfectamente entre ERIC e ISA, eliminando las referencias repetidas; pero en LISA se han producido 5 fallos debido a: 2 errores de mecanografiado en el nombre del autor; en una referencia falta la mitad final del título; en otra referencia falta la mitad inicial del título; y en otra referencia en lugar de «squared» se ha introducido un 2. Parece evidente que estos errores deben evitarse, máxime cuando en las otras bases de datos la coincidencia ha sido plena.

Tabla X
Análisis coste-efectividad de las bases de datos (\$)

	<i>ERIC</i>	<i>LISA</i>	<i>ISA</i>
Coste cita bibliográfica	0,65	1,05	1,15
Total de citas	32	145	80
Coste total	20,8	152,25	92
Efectividad: citas pertinentes	32	128	80
Coste-efectividad por cita pertinente	0,65	1,19	1,15

5 Conclusiones

Como conclusiones, hemos encontrado diferentes argumentos positivos y negativos en las tres bases de datos, destacando lo siguiente:

- ERIC es la base de datos más barata (gratis en Internet), dispone de tesoro y la indización es exhaustiva y precisa, lo que permite altas tasas de precisión y exhaustividad en la recuperación. Sin embargo, son pocas las revistas que vacía en cuanto al tema de «recuperación de la información» en Ciencias de la Documentación (si bien son las más importantes), lo que puede representar un problema en búsquedas exhaustivas sobre el tema, y además el retraso en el tiempo desde que se publica en una revista hasta su incorporación en la base de datos es el más considerable, y por tanto no sirve para una búsqueda actual. Puede servir para una búsqueda complementada con LISA y/o ISA, en un período de más de 1 año, debido a su economía.
- LISA es la base de datos que mayor número de revistas indiza y por consiguiente la que mayor número de publicaciones tiene. El retraso en la actualización es aproximadamente 4 meses para las revistas importantes y mayor para las demás, por lo que tampoco sirve para una búsqueda actual. Es la menos exhaustiva y específica en la indización, lo que puede tener consecuencias negativas en las tasas de exhaustividad y precisión de la recuperación, y como consecuencia de la presencia de referencias no pertinentes (resúmenes de libros), puede resultar la base de datos más cara.
- ISA es una base de datos que cubre un importante número de revistas y además es la única que recoge patentes; dispone de un vocabulario controlado, y la indización es específica y consistente, lo que favorece la precisión y exhaustividad en la recuperación. El tiempo de demora es prácticamente igual que en la base de datos LISA, y resulta ser la más cara por cita bibliográfica recuperada y tiempo de conexión.

Resumiendo, el tiempo de demora en la actualización debería ser mucho menor sobre todo pensando que muchas búsquedas se realizan con el objetivo de conseguir la información más actual; y la política de indización en LISA debería cambiar, tanto en lo que se refiere a especificidad y exhaustividad en la indización como en su política de indizar resúmenes de libros, sobre todo teniendo en cuenta que es una base de datos cara y la aparición de este tipo de referencias encarece la búsqueda.

6 Bibliografía

1. LANCASTER, F. W.; WARNER, A. J. *Information retrieval today*. Washington, D. C.: Information Resources Press, 1993.
2. LANCASTER, F. W. *Vocabulary control for information retrieval*. 2ª edición, Washington, D. C.: Information Resources Press, 1986.
3. CHU, C. M.; AJIFERUKE, I. Quality of indexing in library and information science databases. *Online Review*, 1989, vol. 13, n.º 1, p. 11-28.
4. HARTER, S. P. Semantic relationships between cited and citing articles in library and information science abstracts. *Journal of the American Society for Information Science*, 1993, vol. 44, n.º 9, p. 543-552.
5. SIEVERT, M. E.; VERBECK, A. The indexing of the literature of online searching: a comparison of Eric and Lisa. *Online Review*, 1987, vol. 11, n.º 2, p. 95-104.
6. LABORIE, T.; HALPERIN, M. The Eric and Lisa databases: how the sources of library science literature compare. *Database*, 1984, vol. 4, n.º 3, p. 32-37.

7. BROOKS, T. A. All the rights descriptors: a test of the strategy of unlimited. *Journal of the American Society for Information Science*, 1993, vol. 44, n.º 3, p. 137-147.
8. FENG, S. A comparative study of indexing languages in single and multidatabase searching. *The Canadian Journal of Information Science*, 1989, vol. 14, n.º 2, p. 26-46.
9. HOOD, W.; WILSON, C. S. Indexing terms in Lisa database on CD-ROM. *Information Processing and Management*, 1994, vol. 30, n.º 3, p. 327-342.