

Evaluación del Sistema de Indización SISA

Ana González Olivares

Universidad de Murcia, anagon6@gmail.com

Resumen

La indización es una parte muy importante dentro de la recuperación de información puesto que, cada vez, se hace más difícil dicha recuperación, sobretodo, por el gran volumen de documentos que existen. Se están estudiando desde hace años sistemas automáticos que ayuden al hombre a mejorar la indización. Entre estos sistemas, encontramos a SISA. En este trabajo se hace una evaluación de este sistema, para ello, se realiza una comparación del Sistema de Indización SISA, con la indización manual, empleando los indicadores de consistencia en la indización, exhaustividad y precisión en la recuperación. Aunque SISA admite artículos de cualquier temática, en este trabajo, se ha indizado artículos sacados de la base de datos del ISOC sobre Biblioteconomía y Documentación. Tras el análisis, se ha obtenido un 31% de consistencia en la indización. SISA ha obtenido un 0,8 de exhaustividad y un 0,54 en la precisión y la indización manual, un 0,68 en la exhaustividad y 0,72 de precisión.

Palabras clave: Indización, Indización Automática, SISA, Evaluación de la Indización

1. Introducción

Cada vez se hace más complejo poder encontrar la información que se está buscando, puesto que, existe un gran volumen de documentos a consecuencia de las tecnologías y de las redes, como Internet, que permiten la facilidad de difusión de los documentos, lo que produce que la información se multiplique de forma exponencial. Uno de los conceptos relacionados directamente con este tema, de la recuperación de la información, es la indización, puesto que sin una correcta indización se hace más difícil la correcta recuperación de la información, es decir, la indización nos ayuda a poder encontrar lo que estamos buscando. Generalmente, la indización, la podemos encuadrar dentro del proceso de análisis documental, para Pinto (1993), más concretamente dentro del análisis documental de contenidos. Según Pinto (1993, p.107), se puede decir que la indización es

la técnica de caracterizar el contenido de un documento y/o de las demandas documentales, reteniendo las ideas más representativas para vincularlas a unos términos de indización adecuados, bien procedentes del lenguaje natural empleado por los autores, o de un lenguaje documental previamente seleccionado.

Abstract

Indexing, is a very important part in information retrieval because, every time, is more difficult this retrieval, specially, for the big volume of documents that exists. Since years ago, automatic systems are being studied to help humans to improve indexing. Among this systems, we have SISA. In this paper, an evaluation of SISA is performed. A comparison is made with manual indexing, using for this indicators of consistency in indexing, completeness and precision in the retrieval. While SISA supports articles of any subject, in this paper, articles has been indexed from ISOC database, about Library and Information Science.

After this analysis, a 31% was obtained of consistency in indexing. SISA got 0,8 of completeness and 0.54 of precision, and, manual indexing, 0.68 completeness and 0.72 for precision.

Keywords: Indexing, Automatic Indexing, SISA, Evaluation of Indexing

La indización, generalmente es llevada a cabo por profesionales que se dedican a la realización de este trabajo, pero surgen diversos problemas en la indización manual, como puede ser:

- La gran carga de trabajo, que conlleva que el profesional no pueda dedicarle el tiempo necesario al proceso.
- El error humano, muchas veces consecuencia de la falta de tiempo.
- La intraconsistencia, es decir, que un mismo profesional puede asignar a un mismo documento, distintos términos de indización en dos momentos diferentes, puesto que, en el indizador pueden influir causas que le lleven a esto.
- La interconsistencia, que distintos profesionales pueden proponer distintos términos de indización para un mismo documento.

Todo eso ha hecho que se desarrollen sistemas automáticos que intentan que estos puntos negativos para la indización sean corregidos. Según Stevens (1965), la indización automática "significa el cumplimiento de algunas o todas las operaciones de indización por medios mecánicos".

cos". Por lo tanto, la indización automática es el proceso de asignación de términos que describen el contenido de un documento, para su posterior recuperación, siendo dicho proceso realizado por una máquina sin intervención humana, es decir, los términos escogidos para la indización son los que propone el sistema.

Existen varias ventajas de la indización automática, como la rapidez del proceso o la total consistencia, puesto que estos sistemas, realizan el proceso siempre igual. Se está trabajando e investigado la posibilidad de la creación de sistemas automáticos que realicen este trabajo y que intenten de alguna manera mejorar o al menos igualar la indización realizada por el hombre. Hasta el momento se han desarrollado numerosos sistemas que pretenden imitar la indización realizada por profesionales expertos en la materia. Es importante recordar que existen determinados aspectos que son más efectivos a la hora de realizar la indización, si ésta la realiza un profesional, hay otros, que hacen que sea más efectiva la indización si la lleva a cabo un sistema de indización automática.

Uno de estos sistemas es el analizado en este trabajo: SISA (Sistema de Indización Semiautomática). Si bien su nombre refleja que es un sistema semiautomático, en esta investigación y para su análisis, se trata como un sistema puramente automático, es decir, que en la elección de los términos de indización, no ha habido ninguna intervención humana, se han elegidos solo los propuestos por el sistema. Entre otros objetivos, en este trabajo se pretende realizar una evaluación, a través de casos prácticos y mediante los métodos explicados más adelante, de la indización automática y la indización manual, para comprobar cuáles son los puntos débiles y fuertes del sistema SISA frente a una indización realizada por un profesional.

2. Algunos Sistemas de Indización Automática

Desde finales de los años 50 y principios de los 60 se han desarrollado numerosos proyectos sobre sistemas de indización automática, algunos de ellos solo se han quedado en el papel, otros se aplican a nivel interno en las organizaciones que se han desarrollado y algunos de ellos incluso se han comercializado. Además de estos, existen numerosas investigaciones sobre qué métodos, como por ejemplo, el estadístico, probabilístico, lingüístico, etc., en los que no vamos a entrar en este trabajo, son los más apropiados para ser aplicados por estos sistemas.

Lunh (1959), fué uno de los primeros autores en hablar de la utilización de los ordenadores para la extracción de términos para ayudar a la recuperación de información. Además, según Gil, Rodríguez (1996), este autor habla del método de la frecuencia de aparición de los términos en los documentos. Dice que los términos más adecuados para la indización son los que tienen una frecuencia media de aparición. Fukumoto, F., Suzukit, Fukumoto, J. (1997), en su trabajo, hablan también de la frecuencia de aparición de los términos y proponen un método para extraer palabras claves de artículos dependiendo del contexto de dicho artículo, es decir, que una palabras tienen más importancia que otras, para representar ese artículo, dependiendo del contexto de dicho artículo, es decir, no es lo mismo la palabra "interés" en un artículo sobre "mercado de valores", que en artículo sobre "predicciones meteorológicas".

En Yi-Ming, Pottenger, Schatz (1998), se define el sistema Concept Assigner, aquí se introduce el concepto de Indización semántica, está basado en el Sistema de Redes Neuronales de Hopfield. Según los autores, este sistema atribuye unos conceptos al documento a partir de una colección, sin que estos conceptos aparezcan en el texto del documento. Existen también otros proyectos más actuales como el de Montero (2001), que define el sistema HEPindexer, que es un indexador automático para documentos sobre Física de Altas de Energías y

consiste en el desarrollo de un sistema automático de indexado por asignación. El indexado por asignación consiste en la selección de palabras clave dentro de un léxico controlado.

En Montejo, Dallman (2001), describen un sistema creado para indizar documentos en el CERN (Organización Europea para la Investigación Nuclear), debido al gran volumen de información de este organismos, los profesionales se veían saturados para poder realizar una correcta indización de los documentos, así que se diseñó este sistema para poder ayudarles en el proceso. Worryng, Verster, Todoran, et al. (2001), hablan de un sistema de indización automática para la indización de manuales técnicos. Lo que se pretende en este trabajo es un análisis automático de textos y los gráficos que se encuentran dentro de los manuales técnicos.

3. Sistema de Indización (SISA)

En este apartado vamos a realizar una introducción al Sistema de Indización SISA. Para indagar más sobre el funcionamiento y características de esta aplicación, remitimos a Gil (1997, 1999, 2008).

SISA, como ya hemos dicho anteriormente, es un Sistema de Indización Semiautomático, propuesto en sus inicios para indizar textos correspondientes al área de Biblioteconomía y Documentación. Para realizar el proceso de indización, este sistema, se divide en cuatro módulos (Figura 1.), llamados de preprocesamiento, procesamiento, valoración y el último de presentación de resultados (Gil, 2008, p. 376).

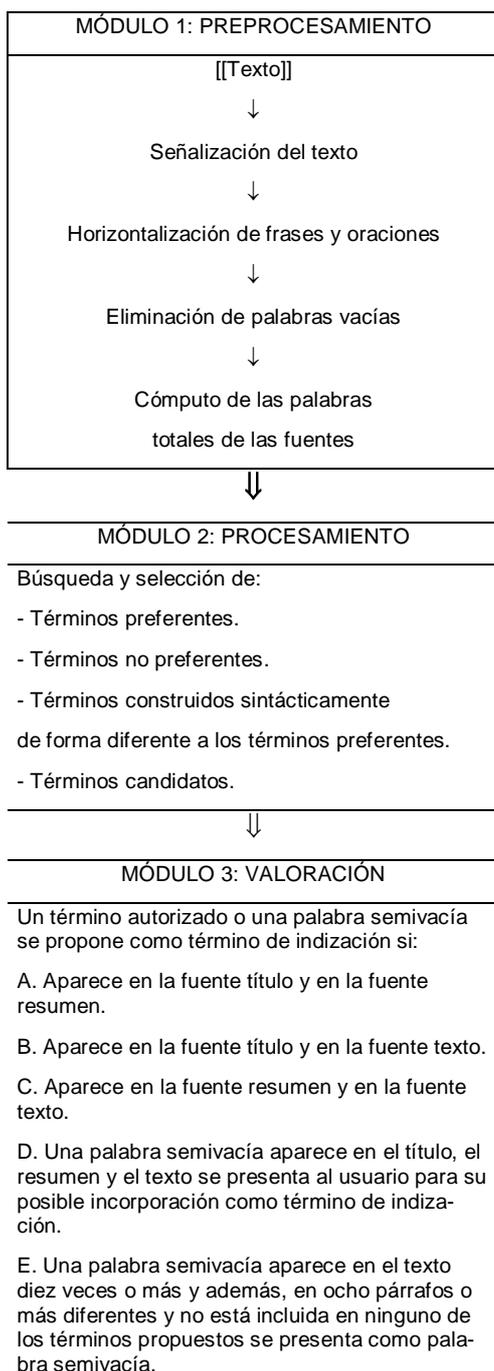


Figura 1. *Arquitectura de SISA*

El sistema utiliza para su procesamiento un vocabulario controlado, donde podemos encontrar términos preferentes y términos no preferentes para la indización, lo que ayuda entre otras cosas a controlar la sinonimia, quizá uno de los factores más difíciles de controlar por los sistemas automáticos.

Este vocabulario relacionado con el tema de Biblioteconomía y Documentación, fue construido a propósito para esta aplicación, a consecuencia, de la falta de vocabularios específicos sobre este tema en castellano. El sistema utiliza este vocabulario para la selección de términos en el texto, mediante un mecanismo de patrones admitidos, es decir, el sistema va comparando los términos que encuentra en el vocabulario, con el texto que se está procesando y selecciona aquellos que coinciden, ya sean preferentes o no preferentes, evidentemente los no preferentes, nos envían a los preferentes.

Además, SISA, realiza una comparación entre términos que coinciden y que se encuentra en el título, resumen y texto completo, y los propone para que el indizador decida si los utiliza o no, esto entra dentro de la definición de sistema semiautomático, por lo que para su evaluación, en este trabajo, no se ha tenido en cuenta.

Para comenzar hay que preparar los textos, estos textos deben estar en formato txt. Para el correcto funcionamiento del sistema, SISA, debe identificar el lugar dentro del documento de donde se han elegido los términos, es decir, si se han elegido de título, resumen, o cuerpo del documento, por lo que, mediante una serie de identificadores que se incorporan manualmente, pero con la ayuda del sistema, que podemos ver más detenidamente en Gil (1999), se le indica al programa el inicio y fin de cada una de estas partes dentro del documento.

Para realizar el análisis de una forma más rápida y para que el sistema no proponga términos sin contenido semántico, la aplicación realiza una eliminación de palabras vacías, es decir, términos como verbos, preposiciones, etc, que no aportan significado por si solas. Esta lista de palabras vacías, es una lista que se incorpora en la configuración del sistema y que se puede ampliar o modificar cuando se quiera.

Tras las dos primeras fases, se pasa al módulo de valoración y ponderación, puesto que no todos los términos del vocabulario que coinciden con los términos del documento son válidos para la indización, se realiza un análisis, para ver qué términos son más adecuados. De manera que se deciden los términos en función

de la parte del documento en la que aparezcan, título y resumen, resumen y cuerpo del documento, título y cuerpo del documento, etc.

Una vez elegidos se presentan al usuario. Por un lado, tenemos los términos seleccionados automáticamente para la indización y por otro la aplicación presenta una serie de términos candidatos, que aunque no están en el vocabulario controlado, tras el módulo de valoración, se considera que pueden ser apropiados para representar el contenido del documento y se da la opción al usuario de incorporarlos, esto hace al sistema semiautomático, la opción que se le da al usuario de participar en la elección de los términos de indización.

La evaluación que hemos realizado en este trabajo, la hemos hecho solo desde el punto de vista automático, es decir, que en la elección de los términos no ha habido intervención humana.

4. Metodología de la evaluación

Para el análisis de SISA se ha realizado una evaluación extrínseca de la indización, es decir, que se han utilizados métodos de comparación de resultados de la indización de dos sistemas, en este caso, una es la indización obtenida con SISA y la otra es la indización humana, realizada por un profesional.

Para la realización de las pruebas se han elegido 30 documentos relacionados con el tema Biblioteconomía y Documentación, obtenidos de forma aleatoria, de la base de datos ISOC. Para la indización de los documentos se ha usado un vocabulario controlado. Los términos del vocabulario están relacionados, con la temática de los documentos. Es importante dejar claro que para el correcto análisis de los resultados, en cualquier evaluación extrínseca, los sistemas a evaluar, deben utilizar un mismo vocabulario controlado. En nuestro caso, cuando se ha incorporado un nuevo término pertinente a la indización en el proceso manual, éste término, se introducía también en el vocabulario controlado de SISA.

Primero, se ha llevado a cabo la indización manual y después se ha realizado la indización con los mismos documentos y con el mismo vocabulario controlado con SISA, para ello se ha utilizado una tabla (Tabla I, en apéndice), donde se pueden ver los títulos de los documentos que se han indizado, los términos de indización con SISA y los términos de la indización manual.

4.1. Evaluación extrínseca mediante la interconsistencia

Existen dos métodos para la realización de la evaluación extrínseca, una mediante la consistencia en la indización y otra mediante la exhaustividad y precisión en la recuperación (Gil, 2008, p. 388)

Para comenzar, con la evaluación primero se ha realizado la evaluación mediante la inconsistencia, una prueba de consistencia de la indización. Para Gil (2002), consistencia "consiste en la búsqueda de las semejanzas durante la asignación de palabras clave, materias o descriptores a un documento", es decir, vamos a realizar una comparación entre las dos sistemas para ver qué términos coinciden entre las dos. Para averiguar este indicador, vamos a utilizar una fórmula de la consistencia que ya ha sido empleada entre otros en Gil (1999, 2002 y 2008), y que es la siguiente:

$$C_i = \frac{T_{co}}{(A + B) - T_{co}}$$

Donde,

C_i = Consistencia entre dos indizadores o sistemas de indización

T_{co} = Número de términos comunes asignados por los dos.

A = Número de términos asignados por A (SISA, en nuestro caso).

B = Número de términos propuestos por el indizador B (Indización manual en nuestro caso)

T_{co} , no siempre corresponde a un número entero, esto ocurre porque hay términos que aunque no son iguales, podrían considerarse casi iguales y se ha hecho una estimación de similitud de 0 a 1, por lo que, por ejemplo, un término asignado por SISA y otro por el profesional podrían tener un T_{co} de 0,5.

La consistencia se ha calculado para cada uno de los artículos indizados, y para averiguar el índice de consistencia final, se ha realizado la media aritmética entre todos los resultados, este resultado se ha multiplicado por 100, para obtenerlo en términos de porcentaje. Esto se podría expresar de la siguiente forma.

$$C_i = \frac{C_{i1} + C_{i2} + C_{i3} \dots C_{in}}{n}$$

Donde,

C_i = Consistencia entre dos indizadores o sistemas de indización

C_i/n = Consistencia de cada uno de los documentos

n = Número total de documentos

El resultado obtenido nos dará el índice de consistencia entre una indización manual y SISA, de acuerdo a la muestra analizada.

4.2. Evaluación extrínseca mediante la recuperación

Una evaluación extrínseca de la indización nos sirve para evaluar la indización basándonos en la recuperación de los documentos. Se puede realizar para evaluar dos indizaciones de un mismo sistema o para evaluar dos sistemas distintos, que es nuestro caso, una indización hecha con SISA y una indización manual hecha por un profesional. Los documentos que hemos utilizado para esta evaluación, son los mismos que se han utilizado para realizar la evaluación extrínseca mediante la interconsistencia, por lo tanto se han aprovechado los términos de indización que se utilizaron para medir la consistencia entre SISA y una indización manual y que podemos ver en (Tabla I, en apéndice).

Para realizar la evaluación extrínseca, mediante la recuperación, hemos creado dos bases de datos, que nos permiten introducir los datos del artículo y los descriptores asignados a cada uno de los artículos indizados. Las dos son idénticas, es decir contienen los mismos campos, los datos son los mismos, a excepción de los descriptores que serán distintos. La base de datos A contendrá los términos de la indización automática mediante SISA, mientras que la base de datos B, contendrá los términos obtenidos de la indización manual. Los campos de las bases de datos serán, identificador del documento, título del documento y términos de indización.

Se han asignado unos temas relevantes para cada uno de los 30 documentos. Cada documento trata sobre unos temas, esto se han establecido para poder saber qué documentos son relevantes y cuáles no (Tabla II, en apéndice).

Después de asignar la relevancia temática a los documentos, se han creado 5 necesidades de información para interrogar la base de datos. Los términos de las preguntas han sido extraídos y posteriormente estos términos se han normalizado mediante el mismo vocabulario controlado que se ha utilizado para la indización de los documentos.

Las preguntas son las siguientes:

1. Documentos sobre recuperación de información en la Web

2. Documentos sobre Archivos o bibliotecas

3. Documentos sobre normalización o metadatos

4. Documentos sobre evaluación de las colecciones

5. Documentos sobre estudios de bases de datos

Con los términos que se obtuvieron de este proceso, se han interrogado las bases de datos, las dos mediante el mismo proceso y con los mismos términos, dándonos unos resultados de la búsqueda. Los resultados se compararon con la tabla de documentos relevantes, para ver las coincidencias. Esto nos permite obtener los datos para poder aplicar las formulas para la evaluación.

Tras interrogar las dos bases de datos con las mismas ecuaciones de búsqueda, con los resultados obtenidos se han aplicado las siguientes fórmulas de exhaustividad y de precisión en la recuperación.

$$\text{Exhaustividad} = \frac{\text{n}^\circ \text{ de documentos relevantes recuperados}}{\text{n}^\circ \text{ total de documentos relevantes en la colección}}$$

$$\text{Precisión} = \frac{\text{n}^\circ \text{ de documentos relevantes recuperados}}{\text{n}^\circ \text{ total de documentos recuperados}}$$

Los resultados obtenidos de aplicar estas fórmulas, deben oscilar entre 0 y 1. Cuanto mayor sea el resultado, es decir, cuanto más cerca esté de 1, mayor será el grado de precisión o de exhaustividad.

A continuación se muestran los resultados obtenidos tras haber efectuado la realización de todo el proceso de evaluación.

5. Resultados

Tras realizar las operaciones que se han llevado a cabo para evaluar el Sistema de Indización SISA, se han obtenido una serie de resultado que comentamos en este apartado.

Además de los índices que se han aplicado para la evaluación de SISA, hemos tomado algunos datos que pensamos pueden ser de interés para realización la evaluación.

Estos datos son, por ejemplo, los tiempos utilizados en la realización del proceso. Los resultados para cada documento mediante SISA y mediante la indización manual, se puede com-

probar en la Tabla I. Se ha calculado el tiempo medio empleado para realizar la indización con SISA 45 segundos y en la indización manual, 24 minutos y 50 segundos. Hay que decir, que en la indización que se ha llevado a cabo mediante SISA, no se ha tenido en cuenta el tiempo empleado para la adaptación de los documentos al formato establecido, pero sí el tiempo empleado en introducir los documentos en el sistema y en la corrección de errores, como la falta de algún identificador en el documento.

ID	SISA	Manual
1	00:20:00	00:00:57
2	00:15:00	00:00:51
3	00:20:00	00:00:35
4	00:31:32	00:00:51
5	00:33:21	00:01:33
6	00:33:09	00:01:55
7	00:30:00	00:00:38
8	00:42:46	00:00:39
9	00:29:40	00:01:55
10	00:28:54	00:01:04
11	00:34:55	00:01:10
12	00:20:07	00:01:15
13	00:17:44	00:00:25
14	00:20:20	00:00:20
15	00:30:29	00:00:24
16	00:29:58	00:00:25
17	00:22:32	00:00:20
18	00:24:51	00:00:20
19	00:27:35	00:00:28
20	00:18:46	00:00:46
21	00:16:23	00:00:34
22	00:15:17	00:00:28
23	00:23:45	00:00:38
24	00:20:08	00:00:26
25	00:27:48	00:00:30
26	00:29:31	00:00:42
27	00:19:11	00:00:39
28	00:13:33	00:00:35
29	00:24:38	00:00:39
30	00:23:20	00:00:42

Tabla I. *Tiempos empleados para la indización*

También se ha efectuado un recuento de los términos de indización empleados, hallando la cantidad de términos totales entre los 30 documentos, una media de términos por documentos y la cantidad de términos simples y compuestos de la indización con SISA y la manual. En la Tabla II se muestran los datos obtenidos.

	Sisa	Manual
Términos Simples	100	65
Términos Compuestos	200	140
Promedio de términos por documento	10	6,8
Total	300	205

Tabla II. *Términos de indización*

5.1. Evaluación extrínseca mediante la interconsistencia

Tras hallar los índices de consistencia de cada uno de los artículos, se efectuó una media aritmética para sacar la media de la consistencia entre la indización automática y la indización manual, el resultado es $C_i=0,31$, es decir, hallamos una consistencia del 31%.

ID	Términos SISA	Términos Manual	Comunes	$C_i = Tco / (A+B) - Tco$
1	2	12	2	0,17
2	11	7	4,5	0,33
3	9	8	3	0,21
4	5	7	0	0
5	11	9	4	0,25
6	7	6	2	0,18
7	9	6	1,5	0,11
8	22	7	2,5	0,09
9	8	6	3	0,27
10	8	6	5	0,55
11	10	6	3	0,23
12	13	7	5	0,33
13	6	4	4	0,67
14	5	8	3	0,3
15	8	9	5,5	0,48
16	9	8	5,5	0,48
17	9	7	4	0,33
18	4	5	1	0,12
19	6	8	4,5	0,47
20	10	6	5	0,45
21	11	6	3,5	0,26
22	8	8	3	0,23
23	9	5	3,5	0,33
24	10	5	3	0,25
25	12	8	6	0,43
26	16	4	3,5	0,21
27	12	7	6	0,46
28	14	7	5	0,31
29	11	5	4,5	0,39
30	14	8	5,5	0,33

Tabla III. *Resultados de los Índices de consistencia*

5.2. Evaluación extrínseca mediante la recuperación

Tras hallar los resultados de consistencia obtenidos, se ha pasado a evaluar la aplicación mediante los índices de exhaustividad y precisión, en función de la recuperación de información.

Para esto, tal y como se ha explicado en la metodología, se ha interrogado una base de datos para obtener los documentos relevantes. Es importante tener en cuenta que la muestra es demasiado pequeña para poder obtener datos más significativos, es aconsejable para este tipo de pruebas utilizar muestras como mínimo de unos 100 documentos.

En la Figura 2, se pueden visualizar los datos que se ha obtenido tras calcular la exhaustividad, para cada una de las preguntas lanzadas a la base de datos, tanto en la indización con SISA, como con la indización manual, obteniendo una media de 0,8 en la primera y 0,69 en el caso de la manual. Por lo tanto, podemos observar como la exhaustividad en SISA, es más alta que en la indización manual.

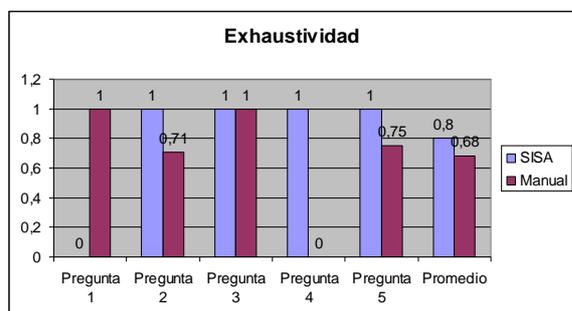


Figura 2. Índices de Exhaustividad

En la Figura 3, se pueden visualizar los datos obtenidos tras calcular la precisión, para cada una de las preguntas lanzadas a la base de datos, tanto en la indización con SISA, como con en la indización manual, obteniendo una media de 0,54 con SISA y 0,72 con la indización manual. Con los datos obtenidos, vemos que la precisión en la indización manual es más alta.

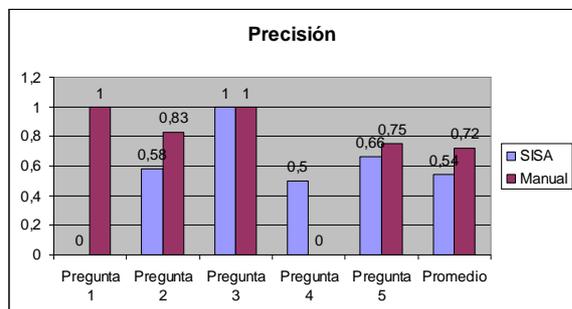


Figura 3. Índices de Precisión

En la Figura 4, se puede ver un gráfico comparativo de los datos de precisión y exhaustividad obtenidos, utilizando la indización realizada con SISA. Podemos ver como el grado de exhaustividad está casi en todos los casos por encima de la precisión.

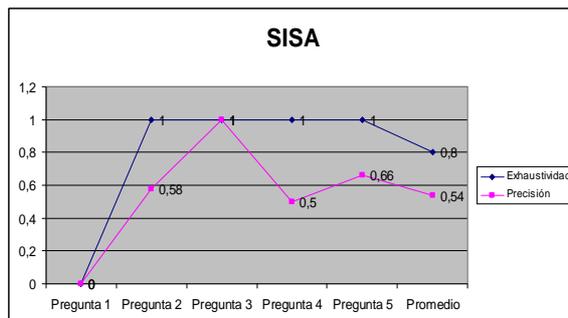


Figura 4. Resultados de SISA

En la Figura 5, se puede ver un gráfico comparativo de los datos de precisión y exhaustividad obtenidos, utilizando la indización manual. Podemos ver como el grado de precisión es ligeramente más alto que el de exhaustividad.

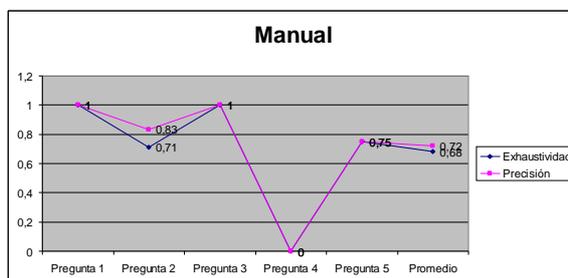


Figura 5. Resultados Indización Manual

En general, los resultados obtenidos han sido muy satisfactorios, incluso más de lo esperado. Los datos de exhaustividad han sido bastante altos, aunque habría que intentar mejorar los datos de precisión para SISA.

6. Conclusiones

El tiempo empleado por el Sistema de Indización SISA, es mucho menor que el empleado por un indizador humano, lo que puede ayudar a realizar una indización más rápida y eficiente, en las empresas u organizaciones que manejen un gran volumen de documentos.

El 31% de consistencia está dentro de los límites que se suelen dar a la hora de medir este indicador, el motivo de esta inconsistencia, es que SISA presenta un número mayor de términos de indización que la indización manual, además de que los términos de SISA son en su

mayoría términos simples, mientras que en la indización manual, se suelen aplicar términos compuestos.

SISA ha obtenido un grado de exhaustividad de 0,8, siendo éste, mayor que en la indización manual donde se ha obtenido un 0,68. Esto nos dice que con SISA se obtienen mayor número de documentos relevantes.

Se confirma que a mayor número de términos propuestos para la indización, mayor será el grado de exhaustividad en la recuperación, puesto que podemos comprobar como el número de términos de SISA es mayor que el número de términos asignados manualmente, y hemos comprobado que la exhaustividad de SISA es mayor que la manual.

En términos de precisión, SISA ha obtenido un resultado de 0,54, por debajo de la indización manual, cuya precisión es de 0,72. Con estos resultados, vemos que la precisión es mayor en la indización manual, por lo que en la indización manual se recuperan menos documentos no relevantes.

La precisión es mayor en la indización manual esto se debe a que aunque el número de términos es menor, los términos que se asignan son más relevantes, además, del uso en la indización humana de los términos compuestos, que son más ricos a nivel de significado, que los términos simples.

Confirmamos la importancia de la indización en la recuperación de la información, puesto que como hemos visto en los casos propuestos, de la indización depende que se obtengan los documentos relevantes para la búsqueda.

Como ya hemos indicado en los resultados, para realizar la evaluación extrínseca mediante la recuperación, se recomienda una muestra de al menos 100 documentos, esto puede dar pie a nuevas investigaciones, para la obtención de datos más significativos.

Referencias

- Fukumoto, F., Suzukit, Y., Fukumoto, J. (1997). An automatic extraction of key paragraphs based on context dependency. *Proceedings of the fifth conference on Applied natural language processing*. 291-298.
- Gil Leiva, I. (1997). *La automatización de la indización: propuesta teórica-metodológica. Aplicación en el área de biblioteconomía y documentación*. (Tesis Doctoral). Universidad de Murcia. Recuperado 06-02-2011 de http://www.tesisenred.net/TDR-0618110-140635/index_cs.html.
- Gil Leiva, I. (2002). Consistencia en la indización de documentos entre indizadores noveles. *Anales de Documentación*, (5), 99-111.

- Gil Leiva, I (1999). *La automatización de la indización de documentos*. Gijón: Trea.
- Gil Leiva, I, (2008). *Manual de Indización: Teoría y práctica*. Gijón: Trea.
- Gil Leiva, I, Rodríguez Muñoz, J. V. (1996). Tendencias en los sistemas de indización automática. Estudio evolutivo. *Revista Española de Documentación Científica*. 19 (3), 273-291.
- Lunh, H.P. (1959). *Keyword in context index for technical literature*. New York: IBM.
- Montejo Ráez, A., Dallman, D. (2001). Experiences in Automatic Keywording of Particle Physics Literature. *High Energy Physics Libraries Webzine*. (5) http://eprints.rclis.org/bitstream/10760/4263/1/Exp_in_Automatic_Keywording_of_Particle_Physics_Literature.pdf.
- Montero Ráez, A. (2001). Proyecto de indexado automático para documentos en el campo de la Física de Altas Energías. *Procesamiento del Lenguaje Natural*. (27), 295-296.
- Pinto Molina, M. (1993). *Análisis documental. Fundamentos y procedimientos*. 2ª ed. Madrid: Eudema
- Stevens, M. E. (1965). *Automatic Indexing: A State-of-the-Art Report*. Washington, DC: National Bureau of Standards.
- Worrying, M., Verster, F., Todoran, L., et al. (2001). Automatic indexing of text and graphics in technical manuals. *International Conference on Multimedia and Expo (ICME)*. 949-952.
- Yi-Ming Chung, Pottenger, W. M., Schatz, B. R. (1998). Automatic Subject Indexing Using an Associative Neural Network. *Proceedings of the third ACM conference on Digital libraries*. 59-68. <http://portal.acm.org/citation.cfm?id=276682>

Apéndice

ID	Título	Términos de indización SISA (A)	Nº términos (A)	Términos de indización manual	Nº términos (B)
1	Conservar sin atesorar: La política de presentación en el Archivo Histórico Eclesiástico de Bizcaia	1.- ARCHIVOS 2.- ARCHIVOS HISTORICOS	2	1.-ARCHIVOS 2.-ARCHIVOS HISTÓRICOS 3.-CONSERVACIÓN 4.-DEPÓSITOS 5.-DIGITALIZACIÓN 6.-DOCUMENTOS ECLESIAÍSTICOS 7.-MANTENIMIENTO DE FONDOS 8.-MICROFILMACIÓN 9.-POLÍTICA DE ARCHIVO 10.-POLÍTICA DE PRESERVACIÓN 11.-PRESERVACIÓN DE DOCUMENTOS 12.-SERVICIO DIOCESANO DE ARCHIVOS (AHEB-BEA)	12
2	El boletín de la Asociación Andaluza de Bibliotecarios: análisis de los artículos de Investigación (1998-2002)	1.-ARTICULOS 2.-ARTICULOS DE INVESTIGACION 3.-ASOCIACION ANDALUZA DE BIBLIOTECARIOS 4.-AUTORIA 5.-BIBLIOTECARIOS 6.-BOLETINES 7.-NVESTIGACION 8.-PAPEL 9.-PROFESIONALES 10.-PUBLICACIONES 11.-PUBLICACIONES PERIODICAS	11	1.-ANÁLISIS BIBLIOMÉTRICOS 2.-ARTÍCULOS CIENTÍFICOS 3.-ARTÍCULOS DE INVESTIGACIÓN 4.-ASOCIACIÓN ANDALUZA DE BIBLIOTECARIOS 5.-BIBLIOTECARIOS 6.-PRODUCCIÓN CIENTÍFICA 7.-PUBLICACIONES PERIÓDICAS	7
3	La situación de las mujeres en las bibliotecas a nivel Internacional	1.-BIBLIOTECAS 2.-CONGRESOS 3.-CORREO ELECTRONICO 4.-ENCUESTAS 5.-IFLA 6.-INVESTIGACION 7.-MESA REDONDA 8.-MUJERES 9.-SALARIOS	9	1.-BIBLIOTECARIAS 2.-BIBLIOTECARIOS 3.-CATEGORÍAS PROFESIONALES 4.-CONDICIONES DE TRABAJO 5.-DIRECTORES DE BIBLIOTECAS 6.-IFLA 7.-MUJERES 8.-SALARIOS	8
4	Las primeras 48 horas Después de un desastre	1.- BIBLIOTECAS 4.- CENTROS 2.- CIENCIAS 3.- COLECCIONES 5.- LIBROS	5	1.-CONSERVACIÓN DE DOCUMENTOS 2.-CONTROL DE PLAGAS 3.-INCENDIOS 4.-INUNDACIONES 5.-PLANIFICACIÓN 6.-PRESERVACIÓN DE DOCUMENTOS 7.-PREVENCIÓN DE RIESGOS	7

5	Estudio bibliométrico de la actividad científica de los matemáticos mexicanos Graduados en Estados Unidos en el Periodo 1980-1998	1.-ACTIVIDADES CIENTIFICAS 2.-BASES DE DATOS 3.-DATOS 4.-ESTADOS UNIDOS 5.-INVESTIGACION 6.-INVESTIGADORES 7.-MATEMATICOS 8.-MEXICO 9.-RESUMENES 10.-SCIENCE CITATION INDEX 11.-UNIVERSIDADES	11	1.-ARTÍCULOS CIENTÍFICOS 2.-BIBLIOMETRÍA 3.-ESTADOS UNIDOS 4.-INDICADORES 5.-INDICADORES BIBLIOMÉTRICOS 6.-INVESTIGADORES 7.-MATEMÁTICOS 8.-MÉXICO 9.-PRODUCCIÓN CIENTÍFICA	9
6	Problemática y tendencias en La arquitectura De metadatos web	1.-FORMATOS 2.-INTEROPERABILIDAD 3.-LENGUAJES 4.-METADATOS 5.-RECURSOS 6.-REGISTROS 7.-WORLD WIDE WEB	7	1.-METADATOS 2.-ONTOLOGÍAS 3.-WORLD WIDE WEB 4.-RDF 5.-RECUPERACIÓN DE LA INFORMACIÓN 6.-WEB SEMÁNTICA	6
7	La edición digital del Boletín Oficial de la Región de Murcia: Parámetros e indicadores Para su evaluación	1.-BOLETIN OFICIAL DE LA REGION DE MURCIA 2.-BOLETINES 3.-DOCUMENTACION 4.-EVALUACION 5.-INDICADORES 6.-PROFESIONALES DE LA INFORMACION 7.-RECURSOS 8.-REGION DE MURCIA 9.-USUARIOS	9	1.- ADMINISTRACIONES PÚBLICAS 2.-BOLETÍN OFICIAL DE LA REGIÓN DE MURCIA 3.-BOLETINES OFICIALES 4.-ESTUDIOS DE USABILIDAD 5.-PÁGINAS WEB 6.-TRATAMIENTO JURÍDICO	6
8	Nuevas conexiones-E-s: la biblioteca Digital Iberoamericana y Caribeña	1.-AUDIO 2.-BASES DE DATOS 3.-BIBLIOTECARIOS 4.-BIBLIOTECAS 5.-BIBLIOTECAS DIGITALES 6.-BIBLIOTECAS NACIONALES 7.-CENTROS 8.-COMUNICACION 9.-DATOS 10.-DIGITALIZACION 11.-DIPLOMADOS 12.-EDUCACION 13.-INVESTIGACION 14.-ORDENADORES 15.-PRODUCTOS 16.-PROYECTOS 17.-RECURSOS 18.-REFERENCIAS BIBLIOGRAFICAS 19.-SERVICIOS DE INFORMACION 20.-SOFTWARE 21.-UNIVERSIDADES 22.-VIDEOS	22	1.-ARCHIVOS ELECTRÓNICOS 2.-BIBLIOTECAS DIGITALES 3.-CARIBE 4.-DIGITALIZACIÓN 5.-IBEROAMÉRICA 6.-RECURSOS ELECTRÓNICOS 7.-TECNOLOGÍAS DE LA INFORMACIÓN	7
9	Las revistas electrónicas: Características, Fuentes de Información y Medios De acceso	1.-DATOS 2.-EMPRESAS 3.-FUENTES DE INFORMACION 4.-INVESTIGACION	8	1.-FUENTES DE INFORMACIÓN 2.-FUENTES DE INFORMACIÓN ELECTRÓNICA	6

		5.-INVESTIGACION CIENTIFICA 6.-PUBLICACIONES 7.-PUBLICACIONES PERIODICAS 8.-REVISTAS ELECTRONICAS		3.-PUBLICACIONES PERIÓDICAS 4.-RECURSOS DE INFORMACIÓN 5.-RECURSOS ELECTRÓNICOS 6.-REVISTAS ELECTRÓNICAS	
10	Biblioteca pública y servicio de información a la comunidad: propuesta para una mayor Integración	1.-ACTUALIDAD 2.-ARTICULOS 3.-BIBLIOTECARIAS 4.-BIBLIOTECAS 5.-BIBLIOTECAS PUBLICAS 6.-INFORMACION LOCAL 7.-SERVICIO DE INFORMACION A LA COMUNIDAD 8.-SERVICIOS DE INFORMACION	8	1.-BIBLIOTECAS 2.-BIBLIOTECAS PÚBLICAS 3.-INFORMACIÓN LOCAL 4.-POLÍTICA DE INFORMACIÓN 5.-SERVICIO DE INFORMACIÓN A LA COMUNIDAD (SIC) 6.-SERVICIOS DE INFORMACIÓN	6
11	Evolución de las bases de Datos jurídicas en España	1.-AUTOMATIZACION 2.-BASES DE DATOS 3.-BASES DE DATOS EN LINEA 4.-BASES DE DATOS JURIDICAS 5.-DATOS 6.-DOCUMENTACION 7.-DOCUMENTACION JURIDICA 8.-MERCADO DE LA INFORMACION 9.-PROYECTOS 10.-WORLD WIDE WEB	10	1.-BASES DE DATOS 2.-BASES DE DATOS JURÍDICAS 3.-DOCUMENTACIÓN JURÍDICA 4.-DOCUMENTOS JURÍDICOS 5.-INFORMACIÓN JURÍDICA 6.-INFORMÁTICA JURÍDICA	6
12	La investigación agrícola en América Latina y el Caribe desde una perspectiva Bibliométrica.	1.-AGRICULTURA 2.-ARTICULOS 3.-CARIBE 4.-CATEGORIAS 5.-CIENCIAS 6.-CITAS 7.-ECONOMIA 8.-IBEROAMERICA 9.-INDICES 10.-INVESTIGACION 11.-INVESTIGACION AGRICOLA 12.-PUBLICACIONES PERIODICAS 13.-SCIENCE CITATION INDEX	13	1.-AGRICULTURA 2.-BIBLIOMETRÍA 3.-CARIBE 4.-IBEROAMÉRICA 5.-INVESTIGACIÓN 6.-INVESTIGACIÓN AGRÍCOLA 7.-PRODUCCIÓN CIENTÍFICA	7
13	La biblioteca pública y la sociedad de la información en Brasil	1.-BIBLIOTECAS 2.-BIBLIOTECAS PUBLICAS 3.-BRASIL 4.-PROYECTOS 5.-SOCIEDAD DE LA INFORMACION 6.- TELECOMUNICACIONES	6	1.-BIBLIOTECAS 2.-BIBLIOTECAS PÚBLICAS 3.-BRASIL 4.-SOCIEDAD DE LA INFORMACIÓN	4
14	Instrumentos de representación del conocimiento: Tesauros	1.- CONVERSION 2.- DOCUMENTACION 3.- ONTOLOGIAS 4.- REPRESENTACION	5	1.-LENGUAJES DOCUMENTALES 2.-ONTOLOGÍAS 3.-ORGANIZACIÓN	8

	Versus ontologías	DEL CONOCIMIENTO 5.- TESAUROS		DEL CONOCIMIENTO 4.-RECUPERACIÓN DE LA INFORMACIÓN 5.-REPRESENTACIÓN DE LA INFORMACIÓN 6.-REPRESENTACIÓN DEL CONOCIMIENTO 7.-TESAUROS 8.-WEB SEMÁNTICA	
15	La formación de usuarios en la biblioteca Pública virtual. Recursos y procedimientos en Las bibliotecas públicas españolas.	1.-BIBLIOTECAS 2.-BIBLIOTECAS PUBLICAS 3.-ESPAÑA 4.-FORMACION DE USUARIOS 5.-RECURSOS 6.-SITIOS WEB 7.-USUARIOS 8.-WORLD WIDE WEB	8	1.-BIBLIOTECAS 2.-BIBLIOTECAS DIGITALES 3.-BIBLIOTECAS PÚBLICAS 4.-EDUCACIÓN A DISTANCIA 5.-ESPAÑA 6.-FORMACIÓN DE USUARIOS 7.-RECURSOS DE INFORMACIÓN 8.-SERVICIOS BIBLIOTECARIOS 9.-WORLD WIDE WEB	9
16	Análisis taxonómico de los portales periodísticos Españoles	1.-COLECCIONES 2.-COMUNICACION 3.-ESPAÑA 4.-INTERNET 5.-MEDIOS DE COMUNICACION SOCIAL 6.-PORTALES 7.-PORTALES PERIODISTICOS 8.-PRODUCTOS 9.-PRODUCTOS INFORMATIVOS	9	1.-ESPAÑA 2.-MEDIOS DE COMUNICACIÓN SOCIAL 3.-PERIÓDICOS ELECTRÓNICOS 4.-PORTALES 5.-PORTALES PERIODÍSTICOS 6.-PRODUCTOS DOCUMENTALES 7.-PRODUCTOS INFORMATIVOS 8.-WORLD WIDE WEB	8
17	Evolución reciente de la Catalogación	1.-BUSQUEDAS 2.-CATALOGACION 3.-NORMALIZACION 4.-RECUPERACION DE LA INFORMACION 5.-REFERENCIAS BIBLIOGRAFICAS 6.-REGISTROS 7.-REGISTROS BIBLIOGRAFICOS 8.-REGLAS DE CATALOGACION 9.-TECNOLOGIAS DE LA INFORMACION	9	1.-CATALOGACIÓN 2.-CATALOGACIÓN AUTOMATIZADA 3.-CONTROL DE AUTORIDADES 4.-METADATOS 5.-NORMALIZACIÓN 6.-REGISTROS BIBLIOGRÁFICOS 7.-REGLAS DE CATALOGACIÓN	7
18	Reflexiones sobre la evaluación de los sistemas de recuperación de información: necesidad, utilidad Y viabilidad	1.- EVALUACION 2.- NECESIDAD DE INFORMACION 3.- SISTEMAS DE RECUPERACION DE INFORMACION 4.- USUARIOS	4	1.-INDIZACIÓN AUTOMÁTICA 2.-PRECISIÓN 3.-RECUPERACIÓN DE LA INFORMACIÓN 4.-RELEVANCIA 5.-SISTEMAS DE RECUPERACIÓN DE INFORMACIÓN	5
19	Evaluación de colecciones en las bibliotecas Universitarias (II).	1.- BIBLIOTECAS 2.- BIBLIOTECAS UNIVERSITARIAS	6	1.-BIBLIOTECAS 2.-BIBLIOTECAS UNIVERSITARIAS	8

	Métodos basados en el uso De la colección	3.- COLECCIONES 4.- EVALUACION 5.- EVALUACION DE COLECCIONES 6.- METODOS DE EVALUACION		3.-COLECCIONES 4.-EVALUACIÓN DE COLECCIONES 5.-EVALUACIÓN DE FONDOS 6.-INDICADORES DE CALIDAD 7.-USO DE LAS COLECCIONES 8.-USO DE LOS FONDOS	
20	El servicio de bibliobús en la Región de Murcia: dos décadas de compromiso Con la igualdad de acceso a la lectura y a la Información	1.-ACCESO A LA INFORMACION 2.-BIBLIOBUSES 3.-BIBLIOTECAS 4.-BIBLIOTECAS MOVILES 5.-COMUNICACION 6.-CONGRESOS 7.-CULTURA 8.-LECTURA 9.-REGION DE MURCIA 10.-SOCIEDAD DE LA INFORMACION	10	1.-ACCESO A LA INFORMACIÓN 2.-BIBLIOBUSES 3.-BIBLIOTECAS MÓVILES 4.-PRÉSTAMO BIBLIOTECARIO 5.-REGIÓN DE MURCIA 6.-SOCIEDAD DE LA INFORMACIÓN	6
21	Evaluación de software De gestión de tesauros	1.-COLECCIONES 2.-EVALUACION 3.-GESTION DE TESAUROS 4.-IMPRESORAS 5.-MANTENIMIENTO DE TESAUROS 6.-MULTITES 7.-SOFTWARE 8.-SOFTWARE DE GESTION DE TESAUROS 9.-STRIDE 10.-TERMINOS 11.-TESAUROS	11	1.-AUTOMATIZACIÓN DE TESAURUS 2.-EVALUACIÓN DE SOFTWARE 3.-GESTIÓN DE TESAUROS 4.-SOFTWARE DE GESTIÓN DE TESAUROS 5.-TESAUROS 6.-TESAUROS AUTOMATIZADOS	6
22	Producción científica de Andalucía en las bases de datos Science Citation Index e Índice Español De Ciencia y Tecnología	1.- BASES DE DATOS 2.- CIENCIAS 3.- CIENCIA Y TECNOLOGIA 4.- DATOS 5.- INDICES 6.- PRODUCCION CIENTIFICA 7.- PUBLICACIONES PERIODICAS 8.- SCIENCE CITATION INDEX	8	1.-ANÁLISIS BIBLIOMÉTRICOS 2.-ANDALUCÍA 3.-BASES DE DATOS BIBLIOGRÁFICAS 4.-BIBLIOMETRÍA 5.-ICYT 6.-ÍNDICE ESPAÑOL DE CIENCIA Y TECNOLOGÍA 7.-PRODUCCIÓN CIENTÍFICA 8.-SCIENCE CITATION INDEX	8
23	Rendimiento de los sistemas de recuperación de información en la web: Evaluación de servicios De búsqueda (Search engines)	1.-BUSCADORES 2.-BUSQUEDAS 3.-CONSULTAS 4.-EVALUACION 5.-EXHAUSTIVIDAD 6.-PRECISION 7.-RELEVANCIA 8.-SISTEMAS DE RECUPERACION DE INFORMACION 9.-WORLD WIDE WEB	9	1.-BUSCADORES 2.-RECUPERACIÓN DE LA INFORMACIÓN 3.-SERVICIOS DE BÚSQUEDA 4.-SISTEMAS DE RECUPERACIÓN DE INFORMACIÓN 5.-WORLD WIDE WEB	5

24	XML como medio de normalización y desarrollo Documental	1.-ARTICULOS 2.-DOCUMENTACION 3.-DOCUMENTALISTAS 4.-NORMALIZACION 5.-NORMAS 6.-PROFESIONALES 7.-SISTEMAS DE INFORMACION 8.-SOFTWARE 9.-WORLD WIDE WEB 10.-XML	10	1.-LENGUAJE HTML 2.-NORMALIZACIÓN 3.-SGML 4.-WORLD WIDE WEB 5.-XML	5
25	Bancos de imágenes y sonido y motores de Indización en la www	1.-ARCHIVOS 2.-BANCO DE IMAGENES 3.-DOCUMENTACION 4.-DOCUMENTACION AUDIOVISUAL 5.-FILMOTECAS 6.-IMAGENES 7.-INDIZACION 8.-INTERNET 9.-MOTORES DE INDIZACION 10.-MUSEOS 11.-SONIDOS 12.-WORLD WIDE WEB	12	1.-BANCO DE IMÁGENES 2.-BASES DE DATOS 3.-BUSCADORES 4.-DOCUMENTACIÓN AUDIOVISUAL 5.-INDIZACIÓN 6.-MOTORES DE INDIZACIÓN 7.-SONIDOS 8.-WORLD WIDE WEB	8
26	Evaluación de la base de datos ISOC a través de un estudio de usuarios. Homenaje a Jose María Sánchez Nistal	1.-BASES DE DATOS 2.-BIBLIOTECARIOS 3.-BIBLIOTECARIOS DE REFERENCIA 4.-CALIDAD 5.-CONSULTAS 6.-DATOS 7.-DOCENTES 8.-ENCUESTAS 9.-ESTADO DE LA CUESTION 10.-ESTUDIO DE USUARIOS 11.-EVALUACION 12.-GRADO DE SATISFACCION 13.-INDICADORES 14.-ISOC 15.-TERMINOS 15.-USUARIOS	16	1.-BASES DE DATOS 2.-ESTUDIO DE USUARIOS 3.-INDICADORES DE CALIDAD 4.-ISOC	4
27	Análisis de uso de las bases de datos de la Biblioteca de la Universidad Carlos III De Madrid	1.-ACCESO A BASES DE DATOS 2.-BASES DE DATOS 3.-BASES DE DATOS EN RED 4.-BIBLIOTECAS 5.-BIBLIOTECAS UNIVERSITARIAS 6.-DATOS 7.-DOCENTES 8.-ESTADISTICAS 9.-INDICADORES 10.-UNIVERSIDAD CARLOS III DE MADRID 11.-UNIVERSIDADES 12.-USUARIOS	12	1.-BASES DE DATOS 2.-BIBLIOTECAS 3.-BIBLIOTECAS UNIVERSITARIAS 4.-ESTUDIO DE USUARIOS 5.-INDICADORES BIBLIOMÉTRICOS 6.-UNIVERSIDAD CARLOS III MADRID 7.-UNIVERSIDADES	7
28	El consumo de información en la industria	1.-ACTUALIDAD 2.-CENTROS	14	1.- CINDOC 2.-CONSUMO DE	7

	farmacéutica de la Comunidad de Madrid. I. Resultados del análisis de la Demanda al Cindoc	3.-CENTROS DE DOCUMENTACION 4.-CINDOC 5.-CLIENTES 6.-COMUNIDAD DE MADRID 7.-CONSUMO DE INFORMACION 8.-DATOS 9.-DEMANDA DE INFORMACIÓN 10.-DOCUMENTACION 11.-DOCUMENTACION CIENTIFICA 12.-EMPRESAS 13.-INDUSTRIA FARMACEUTICA 14.-INVESTIGACION		INFORMACIÓN 3.-DEMANDA DE INFORMACION 4.-ESTUDIO DE USUARIOS 5.-INDICADORES BIBLIOMÉTRICOS 6.-INDUSTRIA FARMACÉUTICA 7.-COMUNIDAD DE MADRID	
29	Elección de puntos de acceso de autor. Implicaciones de las reglas de catalogación en los catálogos automatizados de las Bibliotecas Universitarias españolas	1.-AUTORIA 2.-BIBLIOTECAS 3.-BIBLIOTECAS UNIVERSITARIAS 4.-CATALOGACION 5.-CATALOGOS 6.-CATALOGOS AUTOMATIZADOS 7.-DATOS 8.-INDICES 9.-PUNTOS DE ACCESO 10.-REGISTROS 11.-REGLAS DE CATALOGACION	11	1.-BIBLIOTECAS UNIVERSITARIAS 2.-CATÁLOGOS AUTOMATIZADOS 3.-CATÁLOGOS EN LÍNEA 4.-PUNTOS DE ACCESO 5.-REGLAS DE CATALOGACIÓN	5
30	Las revistas electrónicas en Ciencias Sociales y Humanidades	1.-CALIDAD 2.-CIENCIAS 3.-CIENCIAS SOCIALES 4.-COMITES 5.-EDICION ELECTRONICA 6.-EDITORIALES 7.-EVALUACION 8.-HUMANIDADES 9.-INTERNET 10.-MUESTREOS 11.-PUBLICACIONES PERIODICAS 12.-REVISTAS ELECTRONICAS 13.-TITULOS 14.-UNIVERSIDADES	14	1.-CIENCIAS SOCIALES 2.-EDICIÓN ELECTRÓNICA 3.-EVALUACIÓN DE REVISTAS 4.-HUMANIDADES 5.-INTERNET 6.-RECURSOS ELECTRÓNICOS 7.-REVISTAS CIENTÍFICAS 8.-REVISTAS ELECTRÓNICAS	8

Tabla I. Términos de indización

ID	Títulos	Relevante para...
1	Conservar sin atesorar: La política de presentación en el Archivo Histórico Eclesiástico de Bizcaia	Archivos históricos, Archivos eclesiásticos, Conservación de documentos, Preservación de documentos, Conservación de documentos, Prevención de Riesgos, Archivos
2	El boletín de la Asociación Andaluza de Bibliotecarios: análisis de los artículos de investigación (1998-2002)	Bibliometría, Bibliotecarios, Asociación Andaluza de Bibliotecarios, Artículos
3	La situación de las mujeres en las	Bibliotecas, Mujeres, Bibliotecarias

	bibliotecas a nivel internacional	
4	Las primeras 48 horas después de un desastre	Preservación de documentos, Prevención de riesgos, Bibliotecas
5	Estudio bibliométrico de la actividad científica de los matemáticos mexicanos graduados en Estados Unidos en el periodo 1980-1998	Bibliometría, Estudios bibliométricos, Actividades científicas, Matemáticos, Estados Unidos
6	Problemática y tendencias en la arquitectura de metadatos web	Metadatos, World Wide Web, Recuperación de la información
7	La edición digital del Boletín Oficial de la Región de Murcia: Parámetros e indicadores para su evaluación	Boletín Oficial de la Región de Murcia, Edición electrónica, Evaluación, Indicadores.
8	Nuevas conexión-E-s: la biblioteca Digital Iberoamericana y Caribeña	Bibliotecas digitales, Iberoamerica, Caribe,
9	Las revistas electrónicas: Características, Fuentes de Información y Medios de acceso	Revistas electrónicas, Fuentes de información, publicaciones periódicas
10	Biblioteca pública y servicio de información a la comunidad: propuesta para una mayor integración	Bibliotecas públicas, Servicios de Información a la comunidad, Información Local, bibliotecas
11	Evolución de las bases de datos jurídicas en España	Bases de datos, España, Información jurídica, evolución histórica
12	La investigación agrícola en América Latina y el Caribe desde una perspectiva bibliométrica.	Caribe, bibliometría, Investigación Agrícola, Iberoamerica
13	La biblioteca pública y la sociedad de la información en Brasil	Bibliotecas públicas, Sociendad de la información, Basil, bibliotecas
14	Instrumentos de representación del conocimiento: Tesauros versus ontologías	Tesauros, ontologías, representación del conocimiento
15	La formación de usuarios en la biblioteca pública virtual. Recursos y procedimientos en las bibliotecas públicas españolas.	Formación de usuarios, Bibliotecas públicas, Bibliotecas digitales, España, Recursos
16	Análisis taxonómico de los portales periodísticos españoles	Portales periodísticos, España, periódicos eletrónicos, evaluación
17	Evolución reciente de la catalogación	Evolución histórica, catalogación
18	Reflexiones sobre la evaluación de los sistemas de recuperación de información: necesidad, utilidad y viabilidad	Evaluación, Sistemas de recuperación de información, recuperación de información, necesidad de información
19	Evaluación de colecciones en las bibliotecas universitarias (II). Métodos basados en el uso de la colección	Evaluación, Bibliotecas universitarias, colecciones, bibliotecas
20	El servicio de bibliobús en la Región de Murcia: dos décadas de compromiso con la igualdad de acceso a la lectura y a la información	Región de Murcia, Bibliotecas móviles, acceso a la información
21	Evaluación de software de gestión de tesauros	Evaluación, Software de Gestión de Tesauros
22	Producción científica de Andalucía en las bases de datos Science Citation Index e Índice Español de Ciencia y Tecnología	Producción científica, Andalucía, bibliometría, Bases de datos, Science Citation Index, Índice Español de Ciencia y Tecnología
23	Rendimiento de los sistemas de	Sistemas de recuperación de la

	recuperación de información en la web: Evaluación de servicios de búsqueda (Search engines)	información, recuperación de la información, world wide web, búsqueda de información
24	XML como medio de normalización y desarrollo documental	Metadatos, Normalización , web semántica, XML
25	Bancos de imágenes y sonido y motores de indización en la www	world wide web, Bancos de imágenes, bancos de sonidos, indización automática,
26	Evaluación de la base de datos ISOC a través de un estudio de usuarios. Homenaje a Jose María Sánchez Nistal	Estudios de usuarios, bases de datos, ISOC, Evaluación
27	Análisis de uso de las bases de datos de la Biblioteca de la Universidad Carlos III de Madrid	Bases de datos, bibliotecas universitarias, Universidad Carlos III de Madrid, Comunidad de Madrid, bibliotecas
28	El consumo de información en la industria farmacéutica de la Comunidad de Madrid. I. Resultados del análisis de la demanda al Cindoc	Industria farmacéutica, Comunidad de Madrid, CINDOC, demanda de información, consumo de información
29	Elección de puntos de acceso de autor. Implicaciones de las reglas de catalogación en los catálogos automatizados de las bibliotecas universitarias españolas	Bibliotecas Universitarias, España, Catalogación automatizada, reglas de catalogación, puntos de acceso
30	Revistas electrónicas en ciencias sociales y humanidades	Revistas electrónicas, Ciencias sociales, humanidades

Tabla II. Documentos relevantes