

TRANSPARENCIA N° 1

UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO
CENTRO UNIVERSITARIO DE INVESTIGACIONES BIBLIOTECOLOGICAS

TALLER DE INDIZACION PARA MIEMBROS DE INFOBILA

Ciudad de México, UNAM/CUIB, Septiembre 1998

MARIA TEXIA IGLESIAS MATURANA, M.L.S.

Profesor Titular
Universidad Tecnológica Metropolitana, Santiago, Chile

Fax: 56 2 2250243 Fono: 6777960
e.mail: tiglesia@omega.utem.cl

TRANSPARENCIA N° 2

Este taller estará centrado básicamente en dos PROPOSITOS:

- ①. Revisar algunos principios y metodologías básicas de indización;
- ②. Aplicar estas metodologías analizando documentos reales utilizando el Tesauro Latinomericano de Bibliotecología y Ciencia de las Información, en su versión preliminar.

Por ende: revisaremos los principios y metodologías a través de algunas clases expositivas y luego realizaremos una aplicación real utilizando documentos de la Biblioteca del CUIB y el Tesauro.

TRANSPARENCIA N° 3

EL ANALISIS CONCEPTUAL.

El Análisis Conceptual, o análisis de contenido, consiste en:

* identificar los temas significativos de que trate un documento, lo que nos permite realizar su correcta descripción temática, y, si se desea, una descripción geográfica y/o temporal.

* se trata de identificar posibles puntos de acceso al documento, a partir del contenido temático de éste, expresando este contenido en CONCEPTOS, independiente de las palabras que haya utilizado un autor para expresar su mensaje.



TRANSPARENCIA N° 4

Dentro de este contexto, el Análisis Conceptual, es un PROCESO INTELLECTUAL, a través del cual un documento es analizado para descubrir su real contenido temático.

Insistimos en que se trata de un trabajo intelectual, que permite identificar conceptos, para diferenciar claramente este quehacer de los procesos automáticos de análisis, que dependen en gran medida de las palabras que puedan haber sido usadas para expresarlos.

TRANSPARENCIA Nº 5

El analista enfrenta el siguiente problema:

1. **debe procurar inferir las intenciones del autor cuando escribió su documento (que es una fuente que no está presente);**
2. **debe procurar inferir las necesidades de un potencial usuario que deseará leer el documento (que tampoco está presente).**

TRANSPARENCIA N° 6

El autor suele darle un significado especial, o desea destacar algunas variables o aspectos de su mensaje.

El trabajo del analista es **descubrir** estos valores, a fin de interpretar correctamente el mensaje.

Pero, además, tiene la **responsabilidad de realizar** este trabajo a partir de su propio juicio.

TRANSPARENCIA N° 7

PRODUCTOS DEL ANALISIS CONCEPTUAL,

① RESUMENES

② TERMINOS DE INDIZACION

La indización temática y la redacción de resúmenes, son actividades estrechamente relacionadas, ya que ambas son producto de un trabajo intelectual realizado con la intención de representar el contenido temático de los documentos. ✕

1) **Resúmenes**, el analista prepara una **descripción** narrativa o concisa del documento.

2) **Indización**, el indizador describe el mismo contenido temático, empleando uno o varios términos, generalmente seleccionados de algún tipo de vocabulario controlado y/o esquema de clasificación.

TRANSPARENCIA N° 8

El principal objetivo del resumen es indicar DE QUE TRATA un documento, o sintetizar su contenido. Un conjunto de términos de indización, cumple el mismo objetivo.

Veámos el siguiente ejemplo:

CENTROS DE INFORMACION
RECURSOS COMPARTIDOS
CATALOGOS COLECTIVOS
CATALOGACION COOPERATIVA
REDES EN LINEA
PRESTAMO INTERBIBLIOTECARIO

Este conjunto de términos proporciona una buena idea sobre lo que es tratado en este documento hipotético. Hasta cierto punto, esta lista de términos puede mirarse como un "mini-resumen".

En otras palabras, los términos asignados por el indizador sirven como puntos de acceso para localizar o recuperar un documento, ya sea en un índice impreso o en una base de datos legible por computador.

Dadas las facilidades que ofrecen los sistemas computarizados en línea para realizar búsquedas en texto libre, a mi juicio, la indización y la redacción de resúmenes deben verse como actividades complementarias.

TRANSPARENCIA N° 9

PRINCIPIOS DE INDIZACION.

ETAPAS EN LA INDIZACION.

La Indización Temática es un proceso bi-etápico:

1. Análisis conceptual
2. Traducción al lenguaje de indización

Intelectualmente son etapas completamente distintas. En la práctica no siempre son diferenciadas con claridad y pueden ocurrir simultáneamente, aunque ello no sea recomendable.

TRANSPARENCIA N° 10

A) Análisis conceptual.

El análisis conceptual implica **decidir de qué trata un documento**. Cuál es el tema.

Esta es una afirmación muy simplificada.

La indización temática, por lo general, se realiza para satisfacer las necesidades de determinado público.

* no sólo hay que decidir de qué trata el documento, sino además, por qué podría ser potencialmente útil para determinados usuarios.

* ~~no~~ hay UN conjunto "CORRECTO" de términos de indización para un "determinado item".

* La misma publicación puede ser indizada de una forma diferente en distintas unidades de información.

* Existe una cierta diferencia entre una "indización orientada al documento" y una "indización orientada al usuario".

TRANSPARENCIA N° 11

El indizador debe formularse varias preguntas acerca de un documento:

1. ¿De qué trata?
2. ¿Por qué fue incorporado a nuestra colección?
3. ¿Qué aspectos serían de interés para nuestros usuarios?

TRANSPARENCIA Nº 12

EJEMPLO:

"Informe hipotético publicado por la NASA que trata sobre un vuelo espacial tripulado", indizado en tres diferentes unidades de información, con diferentes intereses y diferentes usuarios.

<u>NASA</u>	<u>Fábrica de caucho</u>	<u>Empresa metalúrgica</u>
Indización exhaustiva cubriendo todos los aspectos en un nivel un tanto genérico.	Nuevos compuestos de caucho sintético.	Nuevas técnicas de soldadura y metales involucrados.
-----	-----	-----
-----	-----	-----
-----	-----	-----
-----	-----	-----
-----	TRAJES	VEHICULOS
-----	ESPACIALES	ESPACIALES

Un documento indizado en 3 unidades distintas

TRANSPARENCIA N° 13

NASA: Como se trata de un informe producido por ellos, al ingresarlo a su propia base de datos, probablemente están interesados en todas las facetas y lo indizarán exhaustivamente procurando cubrir todos sus aspectos.

FABRICA DE CAUCHO: Una parte del informe se refiere a los trajes usados por los astronautas, mencionando algunos compuestos nuevos de caucho sintético. Seguramente utilizarán varios términos muy específicos para indizar los nuevos compuestos, y el término genérico TRAJES ESPACIALES.

EMPRESA METALURGICA: Se menciona, además, una nueva técnica de soldadura desarrollada para combinar ciertas aleaciones en la construcción de un vehículo espacial. Aquí será indizado bajo términos específicos sobre soldaduras y el término genérico VEHICULOS ESPACIALES.

Las tres unidades información, han indizado el documento en forma diferente, respondiendo a sus propios intereses y a los de sus usuarios. y ASI ES COMO DEBE SER.

TRANSPARENCIA N° 14

Esto nos lleva a una **interesante lección**, como dice Wilfrid Lancaster:

"Los indizadores necesitan saber mucho mas que los principios de indización. Deben estar, en especial, enteramente familiarizados con los intereses de la comunidad atendida y las necesidades de información de los miembros de esa comunidad".

TRANSPARENCIA Nº 15

B) Traducción al Lenguaje de Indización.

En esta segunda etapa de la indización temática, se convierte el análisis conceptual en un conjunto de términos de indización.

Hay un esfuerzo por representar la sustancia del análisis conceptual usando términos extraídos de alguna forma de vocabulario controlado.

TRANSPARENCIA N° 16

PRE-COORDINACION.

Indices pre-coordinados = la intersección de clases o conjunción temática, para reflejar temas complejos, se efectúa al momento de construir el término de indización.

* Ejemplos: índices impresos y en tarjetas catalográficas tradicionales.

* características:

1. Es difícil representar la multidimensionalidad de las relaciones entre los términos.
2. Los términos sólo pueden ser listados en una secuencia específica (A, B, B, D, E,), lo cual determina que el primer término sea más importante que el resto, (los subencabezamientos, o subdivisiones, están "modificando" el término llamado principal.)
3. Es muy difícil (por no decir imposible) combinar términos al momento de realizar la búsqueda.

TRANSPARENCIA N° 17

EJEMPLO: "Formación de bibliotecarios en América Latina".

* Se asignan 3 encabezamientos de materia:

EDUCACION, BIBLIOTECARIOS, AMERICA LATINA

* La descripción bibliográfica del documento aparece bajo cada encabezamiento, accediendo al documento por cualquiera de ellos.

* Si un usuario quiere documentos que hablen de EDUCACION y BIBLIOTECARIOS, tiene que mirar lo que hay bajo EDUCACION y luego lo que hay bajo BIBLIOTECARIOS, y tratar de recordar los títulos que coincidiesen bajo ambos encabezamientos.

* Proceso tedioso y muy poco efectivo.

TRANSPARENCIA N° 18

Una forma de mejorar esta situación en los catálogos en fichas es usar un encabezamiento como "subencabezamiento", esto es, los términos son **pre-coordinados** en una sola entrada.

BIBLIOTECARIOS - EDUCACION

BIBLIOTECARIOS - AMERICA LATINA

BIBLIOTECARIOS - EDUCACION - AMERICA LATINA

Utilizando encabezamientos y subencabezamientos de materia, se ha logrado, en cierta medida, la deseada conjunción temática.

Mayor problemas = el PUNTO DE ACCESO es siempre el primer elemento del término de indización.

BIBLIOTECARIOS - EDUCACION - AMERICA LATINA,

el punto de acceso es "BIBLIOTECARIOS" y el término está alfabetizado en la letra "B".

TRANSPARENCIA N° 19

- * Una "solución" podría ser rotar los elementos,
- * Cada clase (que conforman el encabezamiento), tendría la posibilidad de aparecer en primer lugar, como punto de acceso en el índice.

BIBLIOTECARIOS - EDUCACION - AMERICA LATINA
EDUCACION - BIBLIOTECARIOS - AMERICA LATINA
AMERICA LATINA - BIBLIOTECARIOS - EDUCACION
BIBLIOTECARIOS - AMERICA LATINA - EDUCACION
EDUCACION - AMERICA LATINA - BIBLIOTECARIOS
AMERICA LATINA - EDUCACION - BIBLIOTECARIOS

* SOLUCION NADA ATRACTIVA NI FUNCIONAL *

TRANSPARENCIA Nº 20

* Para los índices impresos, se han desarrollado algunos programas computacionales para crear un conjunto de entradas en el índice a partir de una secuencia de términos.

* Ejemplo: el SLIC (Selective Listing in Combination), ideado por J.R. Sharp en 1966.

* Organiza la secuencia de términos en orden alfabético. Esta es la primera entrada al índice. El programa genera, luego el resto de las entradas necesarias, siguiendo dos reglas muy simples:

1. Se mantiene un orden alfabético.
2. Las secuencias redundantes se eliminan

América Latina, Bibliotecarios, Educación
América Latina, Educación, Bibliotecarios
Bibliotecarios, América Latina, Educación
Bibliotecarios, Educación, América Latina
Educación, América Latina, Bibliotecarios
Educación, Bibliotecarios, América Latina
América Latina
Bibliotecarios
Educación

Entradas en un Índice SLIC

TRANSPARENCIA N° 21

CONCLUSION:

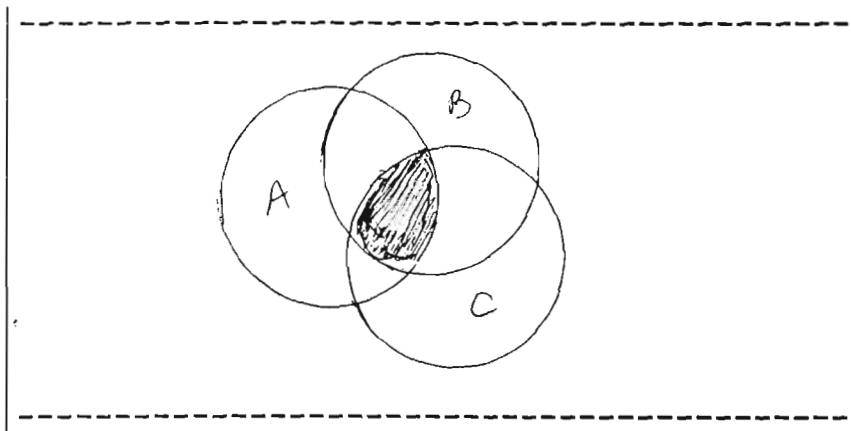
- * Un índice pre-coordinado puede proporcionar muchos puntos de acceso a un documento.
- * Se puede alcanzar un alto nivel de especificidad en el término.
- * Los índices pre-coordinados son rígidos, inflexibles, porque las entradas en el índice deben seguir una secuencia lineal.
- * Las complejas relaciones entre los temas discutidos en los documentos no son unidimensionales, y es muy difícil representarlos con método lineal.

TRANSPARENCIA Nº 22

POST-COORDINACION.

* Una recuperación eficiente, necesita contar con un sistema donde sea factible recuperar un documento por cualquier término o clase, y por cualquier combinación posible de ellos (A, B, D, AB, ABD, BDA, etc.)

* Esta figura intenta graficar la intersección de clases que necesitamos para expresar la multidimensionalidad de los documentos. La zona sombreada muestra la intersección deseada, exactamente.



Intersección de clases

TRANSPARENCIA N° 23

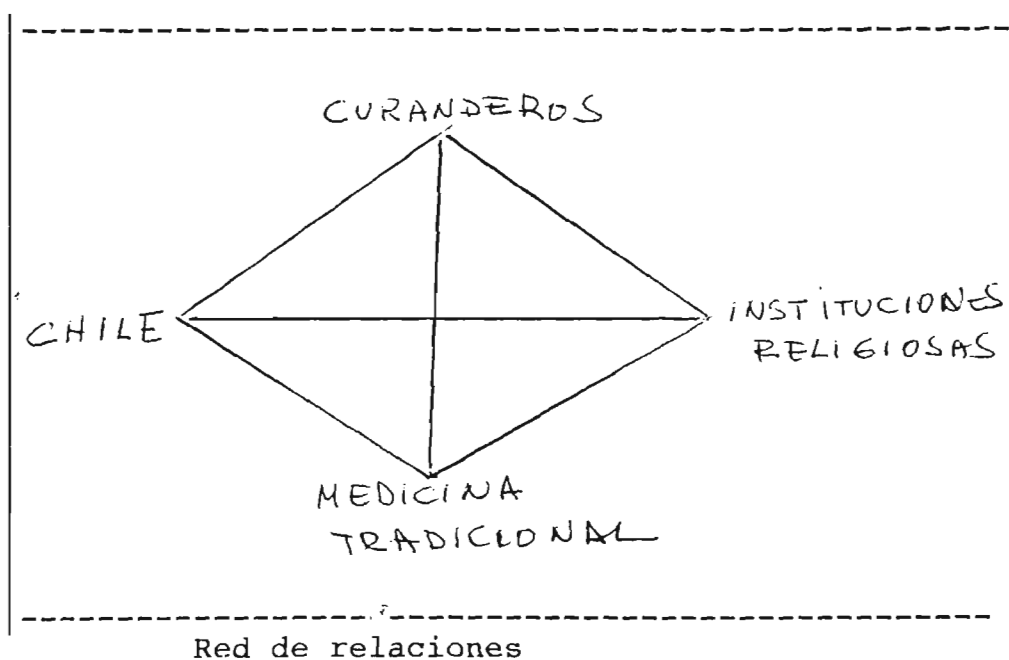
EJEMPLO:

Tema: "La posición de la iglesia frente al trabajo de los curanderos en la zona sur de Chile",

Términos de Indización asignados:

CUARANDEROS
CHILE
INSTITUCIONES RELIGIOSAS
MEDICINA TRADICIONAL

Los términos listados representan una red de relaciones, como muestra el diagrama.



TRANSPARENCIA N° 24

* Debe ser posible recuperar este documento por cualquiera de estos términos o combinación de ellos.

* Un sistema de recuperación que posibilita una búsqueda combinando los términos de cualquier manera, se le conoce como POST-COORDINADO.

TRANSPARENCIA Nº 25

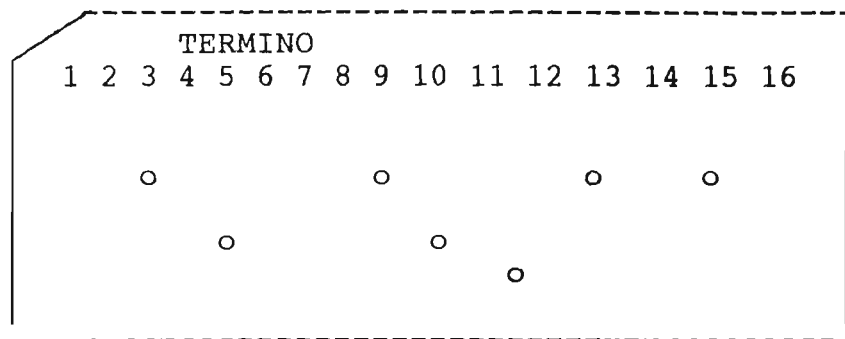
- * Los primeros sistemas verdaderamente post-coordinados aparecieron en los años 40, (Batten en Inglaterra y Cordonier en Francia). Se los conoce como "tarjetas de coincidencia óptica", representado en especial por el sistema "peek-a-boo".

Posteriormente a la invención

TRANSPARENCIA Nº 26

En el Sistema PEEK-A-BOO,

- cada tarjeta representa una clase determinada en el proceso de indización.
- el nombre de la clase, o término de indización, se registra en el borde superior de la tarjeta,
- la superficie restante de la tarjeta está dividida en tal forma que cada cuadrado (presumiendo un arreglo tipo matriz), está dedicado a registrar el número de un documento (o registro) en particular.



Tarjeta Peek-a-boo.

Al indizar el documento del ejemplo, los términos elegidos fueron: "Medicina Tradicional", "Instituciones Religiosas", "Curanderos" y "Chile", y las tarjetas que los representan son perforadas en la posición que corresponda al número identificador del documento.

TRANSPARENCIA Nº 27

El archivo de búsqueda del sistema peek-a-boo consiste en el conjunto de estas tarjetas ordenadas alfabéticamente por el término de indización. Se complementa con un archivo auxiliar de Tarjetas de Registro Bibliográfico, ordenadas secuencialmente por el número del registro.

Descripción bibliográfica	Nº del registro
Resumen (otros datos)	Términos de indización

Tarjeta de Registro Bibliográfico

TRANSPARENCIA N° 28

Para realizar una **búsqueda:**

1° Se decide qué combinación de clases corresponde mas estrechamente al tema que interesa,

2° Se retiran del archivo las tarjetas apropiadas.

3° Las tarjetas se superponen, identificando las posiciones comunes, por las perforaciones que coinciden y dejan pasar la luz.

Como cada perforación corresponde a un número de registro bibliográfico, es posible identificar cuáles documentos han sido indizados bajo los términos que hemos elegido y que, en conjunto, representan el tema que nos interesa.

TRANSPARENCIA Nº 29

* El principio en que se basa el sistema peek-a-boo (y después el Unitérmino), satisface los requerimientos planteado para un sistema de recuperación:

- Se puede buscar por una clase o por cualquier combinación de ellas;
- es igualmente fácil combinar dos clases que varias,
- todas las clases tienen el mismo valor en el sistema;
- no implica secuencia alguna de clases.

* Un sistema computarizado moderno funciona con el mismo principio.

* Un sistema computarizado en línea puede considerarse como un "descendiente directo" de los sistemas post-coordinados manuales.

* Conceptualmente, puede imaginarse, como una matriz:

TRANSPARENCIA Nº 30

términos (clases)	documentos												
	1	2	3	4	5	6	7	8	9	10	11	12	13
A	x					x					x		
B	x	x					x		x				
C				x		x					x		x
D				x	x			x			x		
E	x		x				x			x		x	
F		x	x			x		x					x
G		x			x			x	x	x		x	
H	x			x			x			x			x

Matriz documento - término

TRANSPARENCIA N° 31

Los archivos de un sistema en línea comprenden dos elementos principales:

1. Un conjunto completo de representaciones de documentos (registros bibliográficos), identificados en la matriz por la secuencia de números del 1 al 13.

2. Una lista de términos indicando qué documentos (identificados por el N° de registro) fueron indizados bajo ellos (archivo invertido, "posting file"), identificados en la matriz por la secuencia de letras de la A a la H.

TRANSPARENCIA N° 32

(Taller de Indización T.Iglesias - 1998)

Observando la matriz, podemos visualizar lo que pasa durante una búsqueda en línea.

* Ingresamos el término BIBLIOTECARIOS (representado por la letra B en el diagrama). El sistema responde que hay 4 items indizados bajo ese término.

* Ahora, ingresamos EDUCACION (la letra E en el diagrama), el sistema responde que aparecen 5 items bajo ese término.

* Si luego le pedimos que combine B con E, el sistema comparará los números de registros de las dos listas e indicará que 2 items satisfacen el requerimiento.

* Atendiendo la solicitud del interesado, el computador localiza esos registros por sus números de identificación (1 y 7) y los muestra en la pantalla del monitor o los imprime.

* Este procedimiento permanece igual, no importa cuántos términos o combinación de términos estén involucrados en la búsqueda. o cuántas sean las relaciones lógicas especificadas en la estrategia.

TRANSPARENCIA Nº 25

EN RESUMEN: En los SISTEMAS POST-COORDINADOS,

- 1) Los términos pueden combinarse entre sí de cualquier forma al momento de realizar la búsqueda.
- 2) Se preserva la multidimensionalidad de las relaciones entre los términos.
- 3) Todo término asignado a un documento tiene igual peso - ninguno es mas importante que otro - (a menos que se utilice una indización ponderada).

TRANSPARENCIA Nº 33

* Si la Indización y la Recuperación de *Subversión* son acciones totalmente interdependientes, al indizar un documento para un sistema post-coordinado como la base de datos de INFOBILA, tenemos que pensar en los términos elegidos con la siguiente perspectiva: ellos serán post-coordinados al momento de la búsqueda. X

TRANSPARENCIA N° 34

LA INDIZACION COORDINADA

Si la indización y la recuperación son procesos mutuamente dependientes, en un sistema post-coordinado, no es posible construir una determinada combinación de clases al momento de la búsqueda si no se ha pensado previamente en ella al momento de la indización.

Las relaciones formales entre las clases, y sus posibles operaciones, se manejan con una forma de álgebra reticular llamada álgebra booleana o álgebra de Boole.

* El álgebra de Boole, incluyendo sus generalizaciones respecto a la teoría de conjuntos, es una estructura matemática abstracta simple que muestra estructuras de relación.

* Este último aspecto del álgebra booleana es útil para la indización y para la recuperación de información.

* Las relaciones entre los términos las expresamos como operaciones booleanas.

TRANSPARENCIA N° 35

La teoría de conjuntos es muy útil para comprender mejor las operaciones booleanas, el proceso ^{de} indización, y la distinción entre sistemas pre y post-coordinados.

Un conjunto es una colección de objetos llamados "elementos del conjunto". Puede describirse:

- a) listando todos los elementos del conjuntos;
- b) describiendo una propiedad o característica común a todos los miembros del conjunto.

* En una colección hay una gran variedad de conjuntos de documentos.

* Estos conjuntos han sido descritos a través de propiedades que son comunes a los documentos en la colección.

* La propiedad descrita puede ser el autor, el idioma del texto, el contenido temático, u otra característica.

* Así, un documento puede pertenecer simultáneamente a varios conjuntos diferentes, que corresponden a las propiedades que se reconocen formalmente en un sistema dado.

TRANSPARENCIA N° 36

EJEMPLO: (Extraído de un artículo) de Susan Artandi)

Documento: "la terapia en base a penicilina", publicado
publicado en 1965, por J.Smith.

Puede ser miembro de los siguientes conjuntos:

- CONJUNTO A: (Documentos sobre penicilina)
- CONJUNTO B: (Documentos sobre terapia)
- CONJUNTO C: (Documentos cuyo autor es J. Smith)
- CONJUNTO D: (Documentos publicados en 1965)
- CONJUNTO E: (Documentos publicados en 1965 por J. Smith)
- CONJUNTO F: (Documentos publicados en 1965 por J. Smith
sobre penicilina)
- CONJUNTO G: (Documentos publicados en 1965 por J. Smith
sobre terapia en base a penicilina)
- CONJUNTO H: (Documentos publicados en 1965 sobre uso de
la penicilina)

Distintos conjuntos a que puede pertenecer un
mismo documento.

FUENTE: Artandi, Susan. "Document retrieval and the concept of
sets". Journal of the ASIS. 22(4):289-290, July-
August 1971.

TRANSPARENCIA N° 37

Los **DIAGRAMAS DE VENN** servirán para ilustrar la forma en que se producen las relaciones entre los términos, y visualizar mejor las operaciones propias del álgebra de Boole.

i

* Supongamos una colección de documentos que llamaremos COL y que representaremos con un rectángulo,

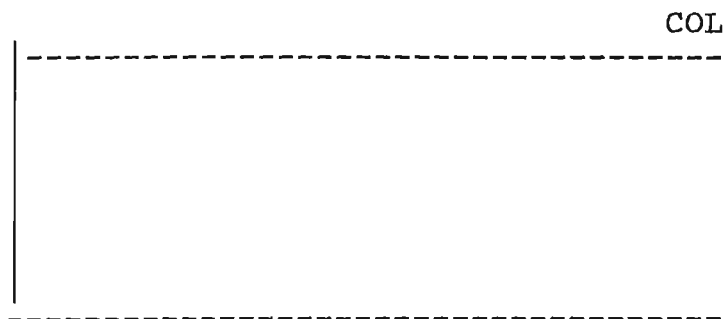


Diagrama de Venn representando una colección de documentos.

i

TRANSPARENCIA N° 38

* Si revisamos uno por uno todos los documentos en COL y a algunos les asignamos el término de indización BIBLIOTECARIOS, hemos creado dos clases:

1. La clase BIBLIOTECARIOS, compuesta por todos los documentos que decidimos **indizar** con ese término, y
2. La clase que llamaremos BIBLIOTECARIOS', formada por los documentos que decidimos **no indizar** con el término BIBLIOTECARIOS.

* La clase BIBLIOTECARIOS' es el **complemento** de BIBLIOTECARIOS y vice versa.

* Una clase no puede tener un complemento a menos que se haya decidido la **pertenencia o no pertenencia** a dicha clase de todos y cada uno de los items en COL.

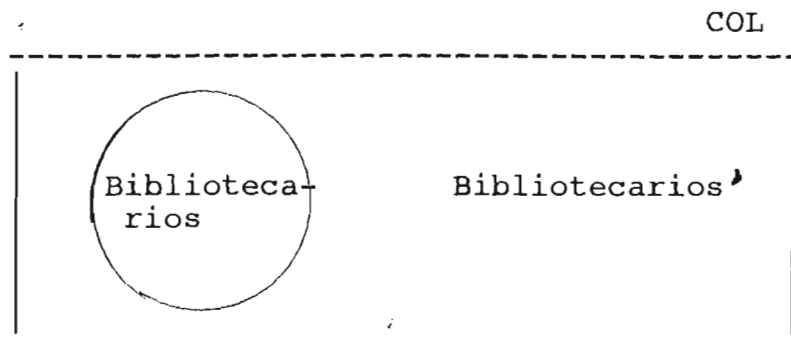


Diagrama de Venn que ilustra la relación de complementariedad.

TRANSPARENCIA N° 39

* La relación entre la clase BIBLIOTECARIOS y la clase BIBLIOTECARIOS' es de complementariedad.

* Esta operación se llama **complementación booleana**, u "Operación Booleana NOT" y se la indica con el símbolo: \neg

* La clase complementaria BIBLIOTECARIOS' es la clase formada por los documentos que, **después de examinados**, no fueron asignados al término BIBLIOTECARIOS.

TRANSPARENCIA N° 40

* Supongamos que formamos otra clase, a la que llamaremos BIBLIOTECOLOGIA, asignando este término de indización a una cantidad de documentos y no asignándolo al resto. Esto crea otras dos clases:

1. La clase de documentos que pertenecen a "Bibliotecarios" o a "Bibliotecología"
2. La clase de documentos que pertenece a ambos, Bibliotecarios y Bibliotecología.

* Esta operación se conoce como "unión", "suma lógica", "operación booleana OR", y se expresa con el símbolo: U

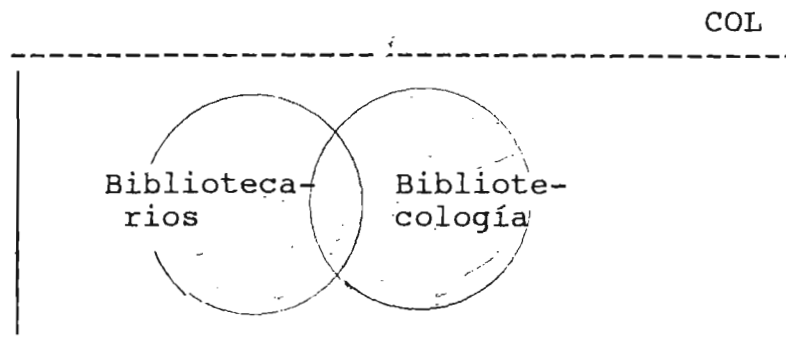


Diagrama de Venn que ilustra la relación de suma lógica.

TRANSPARENCIA Nº 41

* El reconocimiento de esta suma de dos class nos permite, al momento de la búsqueda, pedir documentos que hayan sido indizados por uno u otro término indistintamente.

(BIBLIOTECARIOS U BIBLIOTECOLOGIA)

TRANSPARENCIA Nº 42

* Podemos notar que, además, se ha formado una clase de documentos que pertenecen **simultáneamente** a ambas clases.

* Hay una pertenencia común o conjunta del documento, a Bibliotecarios y a Bibliotecología.

* Esta pertenencia común se conoce como "intersección", "producto lógico", "operación booleana AND" y que se expresa con el símbolo: \cap

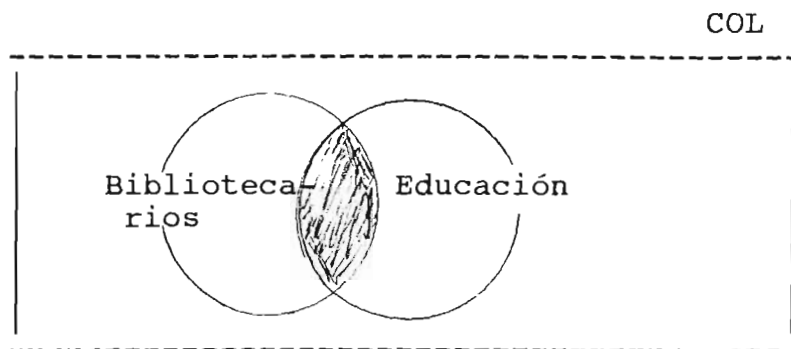


Diagrama de Venn que ilustra la relación de producto lógico.

* El reconocimiento de esta intersección de clases permite, al momento de la búsqueda, solicitar documentos que pertenezcan **simultáneamente** a más de una clase.

* Ejemplo: que pertenezcan a la clase BIBLIOTECARIOS y a la clase EDUCACION. Habiendo sido asignados a ambas clases, se puede presumir que tratarán sobre "Educación de Bibliotecarios". (En la figura la zona sombreada muestra la intersección de clases)

TRANSPARENCIA N° 43

* Es importante **no olvidar** que todas las relaciones de clases descritas tienen validez y significado **si y sólo si** se han tenido en cuenta las condiciones mencionadas:

(a) a todos y cada uno de los items en la colección, definitivamente, se ha asignado o no asignado, todos los términos de indización que correspondan;

(b) estos términos de indización son sólo etiquetas, rótulos; cuánto significado tengan depende de las reglas seguidas para aplicarlos,

* Todas las posibles relaciones entre los términos, apropiadamente definidas, pueden expresarse en términos de:

SUMA DE CLASES	\cup
PRODUCTO DE CLASES	\cap
COMPLEMENTACION DE CLASES	$\bar{}$

La expresión de las relaciones entre los términos suele llamarse "función booleana".

TRANSPARENCIA N° 44

EN RESUMEN:

Podemos decir que los sistemas post-coordinados permiten la coordinación de clases al momento de la búsqueda. Esto permite pedir documentos que hayan sido asignados a la clase A pero no a la B (complementación); documentos que pertenezcan a la clase A o a la clase B (suma lógica); documentos que pertenezcan a la clase A y a la clase B (producto lógico).

TRANSPARENCIA N° 45

EXHAUSTIVIDAD VS. ESPECIFICIDAD.

Hemos dicho que es deber del indizador elegir los conceptos que mejor representen los contenidos de un documento. Para poder elegir correctamente, el indizador necesita guiarse por alguna lógica o pauta de selección.

* Podría guiarse por una cierta categorización reconocida como importante en el área temática cubierta por el documento, (fenómenos, procesos, propiedades, equipos, variables).

* Ejemplo: un documento que trata sobre "uso de la penicilina para tratar determinados cuadros infecciosos". Es probable que el indizador busque **sistemáticamente**, conceptos como "dosis", "efectos secundarios adversos", etc.

Los aspectos mencionados suelen afectar a dos características de la indización: la exhaustividad y la especificidad.

TRANSPARENCIA Nº 46

LA EXHAUSTIVIDAD.

* exhaustividad en la indización, es la medida que todos, o casi todos, los diferentes temas discutidos en un documento, son reconocidos en el proceso de análisis, traducidos al lenguaje del sistema y, por lo tanto, identificados en el índice como términos de indización y, consecuentemente, como puntos de acceso.

EJEMPLO:

Al analizar un documento, el indizador descubre que trata sobre una diversidad de temas complejos,

"Análisis histórico de la trayectoria laboral de egresados de bibliotecología, especialmente aquellos que optaron por las bibliotecas escolares. También analiza brevemente la influencia del Colegio de Bibliotecarios en esta opción. Entrega algunos datos estadísticos respecto a cuántos se quedaron en Santiago y cuántos emigraron a provincias y plantea que de éstos últimos ninguno fué a bibliotecas escolares, prefiriendo las bibliotecas universitarias."

TRANSPARENCIA N° 47

* Supongamos que el indizador reconoce los siguientes temas y los traduce al lenguaje del sistema transformándolos a todos en términos de indización.

- | | |
|-------------------------------|--------------------------------|
| 1. ANALISIS HISTORICO | 7. COLEGIO DE BIBLIOTECARIOS |
| 2. OPORTUNIDAD DE EMPLEO | 8. DATOS ESTADISTICOS |
| 3. OPCION DE TRABAJO | 9. PROVINCIAS |
| 4. BIBLIOTECARIOS | 10. BIBLIOTECAS UNIVERSITARIAS |
| 5. BIBLIOTECAS ESCOLARES | 11. CHILE |
| 6. ESCUELA DE BIBLIOTECOLOGIA | |

* Sin duda se puede decir que el indizador ha sido altamente exhaustivo en la indización de este documento.

* Si todos estos temas son indizados, al momento de la búsqueda el documento puede ser recuperado por cualquiera de estas clases o combinación de ellas.

TRANSPARENCIA Nº 48

* Esta recuperabilidad del documento permite decir que un alto nivel de exhaustividad, asegura un alto nivel de acierto o "recall".

ACIERTO = Capacidad del sistema para recuperar documentos relevantes para una búsqueda en particular, en relación al total de documentos relevantes existentes en la colección.

* Pero, el alto nivel de exhaustividad, si bien favorece el "acierto", puede afectar negativamente la precisión de la recuperación en el momento de la búsqueda.

PRECISION = Capacidad del sistema para recuperar documentos relevantes a una búsqueda en particular, en relación al total de documentos recuperados. Es la capacidad de retener lo no relevante. (Prueba la capacidad de filtro de un índice)

TRANSPARENCIA N° 49

El efecto negativo de la exhaustividad sobre la **precisión**, se debe, básicamente por **dos razones**:

(a) Si para cada documento ingresado se reconocen todos o casi todos los diferentes temas tratados, ese documento será recuperado por cualquiera de esos puntos de acceso (favoreciendo el acierto). Sin embargo, es probable que el documento haya sido indizado bajo temas que, si bien aparecen en el documento, estén tratados superficial o tangencialmente, pero igual el documento será recuperado en respuesta a preguntas sobre cualquiera de estos temas o combinación de ellos.

(b) Mientras más temas se reconozcan en la indización y más términos deban ser usados para expresar o identificar esos temas, mayor es la probabilidad de que se produzca una **falsa coordinación** entre los términos al momento de la búsqueda,

TRANSPARENCIA N° 50

* Ejemplos de preguntas que muestran cómo el documento puede ser recuperado erróneamente:

1) Pregunta: "Historia del Colegio de Bibliotecarios de Chile"

Estrategia: = HISTORIA o ANALISIS HISTORICO
y COLEGIO DE BIBLIOTECARIOS
y CHILE

2) Pregunta: "Mercado laboral en provincias"

Estrategia: = OPORTUNIDAD DE EMPLEO o MERCADO DEL TRABAJO
y PROVINCIAS

3) Pregunta: "Cifras sobre bibliotecarios chilenos"

Estrategia: = BIBLIOTECARIOS y DATOS ESTADISTICOS y CHILE

4) Pregunta: "Datos estadísticos de bibliotecas de educación superior en Chile"

Estrategia: = BIBLIOTECAS UNIVERSITARIAS y CHILE y DATOS
ESTADISTICOS



TRANSPARENCIA N° 51

Dos cosas importantes que es necesario recordar:

(a) La exhaustividad en la indización es una decisión, una política de la unidad de información. Puede ser una política muy positiva en determinadas circunstancias, propias de una situación en particular, pero también riesgosa.

(b) Hay una relación inversa entre el nivel de exhaustividad en la indización y el nivel de precisión en la recuperación.

TRANSPARENCIA Nº 52

ESPECIFICIDAD.

* A diferencia de la exhaustividad, que es una política de indización, la **especificidad en la indización depende del lenguaje de indización.**

* La **ESPECIFICIDAD** consiste en la capacidad del lenguaje de indización utilizado para describir en forma específica, precisa, los diferentes temas tratados en los documentos. X

* **EJEMPLO:**

TEMA: "Indización en cadena":

- En INFOBILA, usando el tesoro especializado, el documento podría ser indizado bajo: INDIZACION EN CADENA.

- En una biblioteca universitaria, pero que no usa este Tesoro, podría indizarse bajo: INDIZACION.

- En una Biblioteca general (si el documento estuviese allí), tal vez estaría indizado bajo BIBLIOTECOLOGIA.

* Al momento de la búsqueda, la situación será diferente en las tres unidades: el nivel de precisión ira decreciendo, pero el nivel de acierto puede subir.

TRANSPARENCIA N° 53

EN RESUMEN:

* la exhaustividad incide en la capacidad de "acierto" de un índice;

* la especificidad de un lenguaje de indización incide en la capacidad de precisión de un índice.

TRANSPARENCIA N° 54

EL PROCESO DE INDIZACION

PASOS EN LA INDIZACION DE UN DOCUMENTO.

El siguiente **diagrama de flujo** muestra una secuencia de los pasos que usualmente se dan, en mayor o menor detalle, en la indización de un documento.

TRANSPARENCIA Nº 55

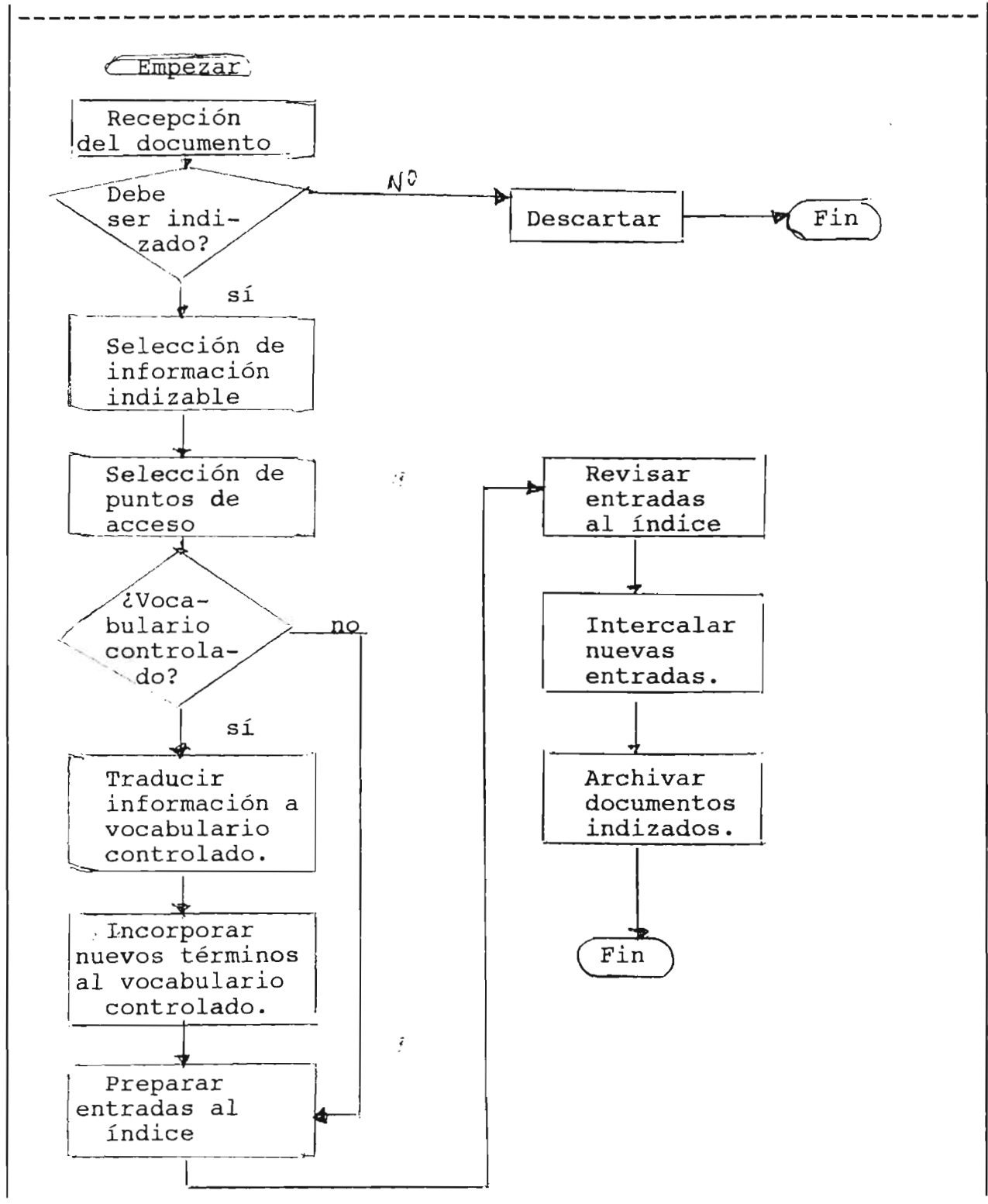


Diagrama de flujo de los Pasos Básicos en la Indización de un documento.

TRANSPARENCIA Nº 56

**ALGUNOS CONSEJOS DEL SISTEMA MEDLARS PARA LA SELECCION DE LOS
TERMINOS CANDIDATOS.**

Las instrucciones son las siguientes:

1. Lea y comprenda el título.
2. Lea el texto hasta encontrar el punto donde el autor establece el propósito del documento.
3. Revise (scan) el texto, leyendo los encabezamientos de capítulos y/o secciones, frases en negrita o en letra cursiva, gráficos, láminas, etc.
4. Lea cada palabra del sumario o tabla de contenido.
5. Revise cuidadosamente el resumen.
6. Revise las citas bibliográficas.

Preguntas de un Revisor.

1. ¿Los términos principales reflejan el verdadero contenido del documento?
2. ¿Están los términos correctamente escritos, de acuerdo al Tesauro?
3. ¿Para este artículo, están los términos al nivel correcto de especificidad?

(Taller de Indización - T.Iglesias - 1998)

TRANSPARENCIA Nº 57

CARACTERISTICAS DEL INDIZADOR,

- 1) Conocimiento del tema.
- 2) Conocimiento acabado del lenguaje de indización que se utiliza.
- 3) Conocimiento de las técnicas y procedimientos de recuperación de información que utiliza el sistema.
- 4) Conocimiento de los requerimientos y características del usuario.

Santiago, Chile, Septiembre de 1998.