

mente y, por otra parte, se muestran explícitamente las relaciones jerárquicas entre los términos, además de las relaciones semánticas y de asociación. A los términos de indización incluidos en un tesoro suelen ser llamarseles **descriptores**. (Ver figura 16).

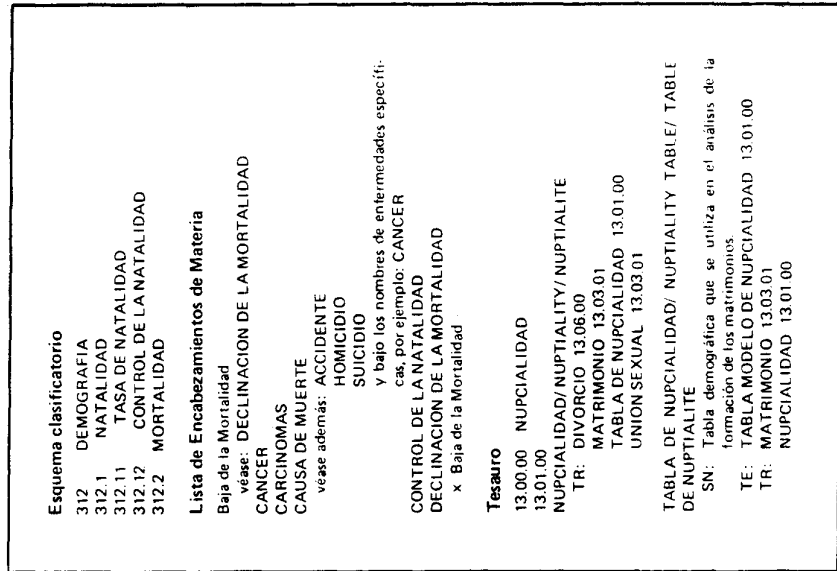


Fig. 16: Términos de indización en diferentes tipos de vocabularios controlados

6. LOS INDICES.

Sin índices, difícilmente podríamos telefonar, ordenar comidas en un restaurant, hacer arreglos para un viaje, encontrar documentos en una biblioteca, calendarizar reuniones, encontrar una calle, y tantas otras cosas que damos por hechas. La interacción social, tal como la conocemos hoy en día, sería imposible sin índices. No se trata de que en otros tiempos no haya habido índices, de seguro los ha habido desde antiguo, pero no en la forma que hoy los conocemos, ni con los métodos de indización que hoy usamos.

Como decía Robert Collison (1), el costurero de la época victoriana con sus cientos de pequeños compartimentos para los diferentes tipos de hilos, agujas, dedales, cintas, tijeras, etc., sin duda era una suerte de índice, sólo que su uso dependía de un enfoque visual y no de una lista escrita de su contenido. El índice existía, no como una serie de palabras, sino como un patrón en la mente del usuario.

Los índices han mejorado extraordinariamente a través de los años. Actualmente, son más acuciosos, más fáciles de usar, y los hay de todo tipo. Además, hoy se almacenan en archivos computarizados y las búsquedas se realizan correlacionando los encabezamientos a través de operadores booleanos. Sin duda el computador ha sido un aporte inapreciable y que ha permitido grandes avances en el campo del análisis y la recuperación de información.

(1) COLLISON, Robert L. *Indexes and Indexing*. London, Ernest Benn Ltd., 1959. p. 13.

Los índices son excelentes guías. No sólo nos llevan a un tema de interés, sino que, además, nos proporcionan un cuadro integrado, una visión global del campo temático, como ninguna otra herramienta bibliográfica puede hacerlo.

6.1. DEFINICIONES.

Hay dos conceptos fundamentales que debemos definir aquí: índice y entrada.

6.1.1. INDICE.

De acuerdo a la Norma ANSI (1), un índice es "una guía sistemática a los items contenidos en, o conceptos derivados de una colección (de documentos). Estos items o conceptos derivados de, están representados por entradas organizadas en un orden de búsqueda conocido o establecido, ya sea alfabético, cronológico o numérico".

Por otra parte, un índice analítico ("back of the book index") es "una lista alfabética de nombres, lugares y/o temas incluidos en un trabajo impreso, el cual proporciona para cada ítem la página en la cual se encuentra". (2)

6.1.2. ENTRADA.

La entrada es el componente principal de un índice. Consiste en: un encabezamiento (palabra, frase o símbolo usado para identificar un nombre, tema, lugar, etc.) y una cierta cantidad de información, la que puede ir desde una descripción bibliográfica completa, hasta el número de una página o un número de acceso.

(1) ANSI Z39.4-1968. Basic criteria for indexes.
(2) AMERICAN Heritage Dictionary.

Dentro del contexto de un "índice analítico" o "índice de una obra", la entrada consiste en un encabezamiento, (por lo general una palabra o frase), una modificación o sub-encabezamiento, y la referencia a la página correspondiente.

6.2. TIPOS DE INDICES.

Los índices pueden clasificarse desde diferentes puntos de vista:

1) Desde el punto de vista físico.

- impresos
- en tarjetas
- en microformas
- en medios magnéticos

2) Desde el punto de vista del arreglo:

- alfabético
- cronológico
- numérico, seriado o secuencial
- ejemplo: Primer Informe Anual
Segundo Informe Anual
- clasificado
- ejemplo: Caminos
Caminos Interprovinciales

3) Desde el punto de vista del elemento ordenador o de entrada:

- de autor
- de título
- de materia
- de conferencias
- etc.

4) Desde el punto de vista estructural:

- pre-coordinados
- post-coordinados

6.3. EL INDICE DE AUTOR.

El **índice de autor** es una guía a los autores de los diversos items (documentos, publicaciones, registros, etc.) contenidos en una colección, bibliografía, antología, publicación periódica, etc., que se esté indizando.

Un **índice de nombres u onomástico**, en cambio, es un índice de los nombres de personas que aparecen nombradas, citadas o analizadas en un documento, que es el tipo de información que suele aparecer en un índice analítico o de una obra, pudiendo también constituir un índice aparte.

Un índice de autor sirve esencialmente para: 1) ubicar un documento en particular del cual se conoce su autor, y su título, y 2) para saber qué trabajos de un determinado autor existen en una colección. Por otra parte, indirectamente, también guían hacia un tema o materia, dado que los escritos de un autor en particular suelen estar relacionados con la materia de su especialidad.

Un **índice de autor simple**, es aquél cuyas entradas consisten en el nombre de los autores, seguidos de bloques de números de referencia. Estos números pueden ser: el número de la cita en el cuerpo principal de una bibliografía o de una revista de índices y resúmenes; el número de registro bibliográfico en una base de datos; el número de acceso del documento en la estantería; o el número de la página en una antología. (ver figura 17)

BEHM, Hugo	268, 269, 315
BELL, Daniel	17, 36
BELTRAN, René	320
BOLTON, Lina	25, 76, 150, 151

Fig. 17: Índice de autor simple.

Estos índices "simples" (ya sean de autor, tema u otro) son económicos, y fáciles y rápidos de preparar, pero su gran desventaja es que, como sólo entregan bloques de números, obligan al lector a revisar una cantidad de referencias hasta encontrar el o los trabajos deseados o relevantes. Así, la ventaja es para el sistema, pero la desventaja para el usuario.

Un elemento modificador que se puede agregar a estos índices y que mejora substancialmente su utilidad, es el título del documento y, eventualmente, la fecha de éste, además de, obviamente, el número o código del documento. (ver figura 18).

CAMPAÑA, Paulo	
41-660	La política migratoria.
41-662	La migración laboral en América Latina.
CARRON, Juan	
41-181	El papel redistributivo del Estado.
41-183	Cuestionario de la encuesta de empleo utilizada en Santiago.

Fig. 18: Índice de autor al que se ha agregado el elemento título.

Naturalmente, un índice de autor incluye todos los **coautores**, cada uno de los cuales debe tener la oportunidad de aparecer como encabezamiento de la entrada. Cabe recordar que, de hecho, en una obra colectiva todos son coautores.

Dependiendo de la política del sistema, puede registrarse sólo un apellido y el nombre, o un apellido y la inicial del nombre, o todos los apellidos y todos los nombres. Dos autores con idéntico nombre pueden diferenciarse con la fecha de nacimiento o algún otro dato por el estilo.

En un índice impreso, el nombre del autor aparece una sola vez y luego bajo él se listan los documentos, ya sea alfabéticamente por el título, secuencialmente por el número de la referencia, o cronológicamente por la fecha de publicación del documento.

Un índice de autor debe también consignar la responsabilidad en relación al documento Y, si no se trata del autor propiamente tal, indicar si es un editor, compilador, traductor, ilustrador, etc.

Un índice de autor puede incluir sólo autores personales, sólo institucionales, o ambos. Es importante que este tipo de índices incluya las correspondientes referencias "véase" para diferentes formas del nombre, pseudónimos, etc.

Naturalmente, todos los aspectos formales que se han mencionado, son válidos para otros índices en que el punto de acceso o elemento principal de la entrada no es el autor sino otro elemento del registro bibliográfico, como la materia, cobertura geográfica, fecha de publicación, etc.

6.4. INDICE TEMATICO.

En este punto parece necesario hacer notar una distinción entre algunos tipos de índices que, teniendo similitudes, también tienen diferencias muy importantes.

1) **Índice de tópicos o Índice de Categorías Temáticas.** Estos "índices" son más bien un arreglo u ordenación que suele darse a una bibliografía, boletín bibliográfico, o una revista de resúmenes. Son grandes categorías que sirven para organizar el cuerpo principal de una publicación, a fin de que el usuario encuentre reunidos los documentos de su área de interés o especialización, o para realizar búsquedas más genéricas en una base de datos.

2) **Índice de palabras.** Los índices KWIC y KWOC son ejemplos de índices de palabras, los que, como ya hemos dicho, no son

índices temáticos en el estricto sentido de la palabra. De hecho se basan en un aspecto circunstancial, esto es, las palabras que aparecen en los títulos, las que pueden ser o no representativas del contenido o tema del documento.

3) El **Índice temático**, en cambio, es aquel índice, producto de un proceso intelectual de análisis, donde los encabezamientos de las entradas están formados por **términos escogidos**, considerados los más representativos del contenido temático de un documento y que, por lo general, han sido seleccionados de un vocabulario controlado.

En el caso de un **índice temático clasificado**, la organización de las entradas sigue el orden de un esquema de clasificación. (ver figura 19). La descripción del documento y su número de referencia aparecen bajo la notación representativa de la clase a que fue asignado el documento. Estos índices se complementan con un **índice alfabético auxiliar** que envía a las clases correspondientes.

621.471
• Principios generales de termodinámica (078)
• Radiación (127)
621.477
• Construcción de hornos solares (530)

Fig. 19: Índice temático clasificado

En el caso de un **índice temático alfabético**, la organización de las entradas obedece a palabras o frases, que son los términos de indización, ordenadas en una secuencia alfabética. Bajo el término aparecen las descripciones de los documentos asignados a la clase representada por ese término.

Los **índices temáticos** suelen incluir un sistema sindético para su mejor uso. El **sistema sindético** (*) de un índice es todo el conjunto de guías utilizadas para indicar conexiones y relaciones entre las entradas, a fin de ayudar al usuario a lograr una mejor recuperación de información.

La guía que proporciona el sistema sindético de un índice incluye: referencias cruzadas, inversiones (con comas o paréntesis) y notas de alcance. La guía está dirigida a: 1) la elección de términos entre sinónimos; 2) encabezamientos relacionados; 3) qué ha sido indizado bajo un encabezamiento; 4) cómo está organizado el encabezamiento; 5) cobertura, alcance, significado y excepciones.

El propósito de una guía sindética es prevenir al usuario sobre entradas que no existen y ahorrar tiempo en la búsqueda. Sin esta guía, el indizador no podría evitar el **dispersar** parecidas e incluso idénticas entradas. El sistema sindético de un índice es planeado, diseñado y construido por el indizador.

6.5. LOS INDICES PRE-COORDINADOS.

Los índices pre-coordinados son aquellos en que la intersección de clases o conjunción temática, necesaria para reflejar temas complejos, se realiza al momento de construir el **término de indización**.

Estos índices pueden ser alfabéticos o clasificados, impresos o en tarjetas.

Veamos el siguiente documento: "Efectos sociales de la drogadicción en adolescentes".

(*) 1. SISTEMA SINDE TICO (syndetic system) = sistema conectante; sistema conjuntivo. 2. SINDE TICO = adjetivo que se aplica a un índice o catálogo que incluye referencias cruzadas de enlace entre los términos. 3. CATALOGO SINDE TICO DE MATHIAS = catálogo cuyas entradas se relacionan por medio de referencias.

(Fleisher's Dictionary of Library Sciences, Information and Documentation, Amsterdam, 1973.)

a) Esquema de clasificación (Dewey):

362.2933 (362.293: drogadicción, 3: efectos sociales)
362.796 (problemas sociales del adolescente)

b) Encabezamientos de materia:

DROGADICCION — ADOLESCENTES — ASPECTOS
SOCIALES

c) Tesauro:

ADOLESCENTE
ESTUPEFACIENTE
PROBLEMA SOCIAL o CONSECUENCIAS SOCIALES

Al momento de la búsqueda en un índice pre-coordinado, como son los catálogos tradicionales de las bibliotecas o los índices impresos, es necesario revisar todas las entradas que aparecen bajo cada uno de estos términos, a fin de estar razonablemente seguros de que se han ubicado todos los documentos que existen en la colección sobre este tema.

La combinación de encabezamientos y sub-encabezamientos de materia, se utiliza con el propósito de lograr la deseada conjunción temática.

Con la notación clasificatoria del ejemplo entregado, se intenta algo parecido. Pero, vemos que aquí sólo aparece la intersección de dos clases, en circunstancias que el documento muestra la intersección de tres temas o clases: drogadicción, adolescentes, problemas sociales.

362.2933 = drogadicción + efectos sociales

362.796 = adolescente + problemas sociales

Sin embargo, como se puede observar, aún cuando se ha logrado la conjunción de dos clases, sólo es posible recuperar el documento si se busca bajo esta determinada notación.

Con los encabezamientos de materia ocurre lo mismo que con las notaciones clasificatorias: el punto de acceso está constituido

por el primer elemento del término de indización. En el caso del ejemplo planteado: DROGADICCIÓN-ADOLESCENTES-PROBLEMAS SOCIALES, el punto de acceso es "Drogadicción" y el término estará alfabetizado en la letra "D".

Una solución en la que podría pensarse, en términos hipotéticos, es **rotar los elementos**, vale decir, que cada una de las clases que conforman este encabezamiento, tenga la posibilidad de aparecer en primer lugar, como punto de acceso en el índice.

DROGADICCIÓN-ADOLESCENTE-PROBLEMAS SOCIALES
 DROGADICCIÓN-PROBLEMAS SOCIALES-ADOLESCENTE
 ADOLESCENTE-DROGADICCIÓN-PROBLEMAS SOCIALES
 ADOLESCENTE-PROBLEMAS SOCIALES-DROGADICCIÓN
 PROBLEMAS SOCIALES-ADOLESCENTE-DROGADICCIÓN
 PROBLEMAS SOCIALES-DROGADICCIÓN-ADOLESCENTE

Es evidente que esta alternativa de solución no es aconsejable, pues tendríamos monstruosos catálogos totalmente inmanejables. Como la conjunción de clases se establece al momento de crear el encabezamiento, sería necesario crear seis encabezamientos diferentes a fin de darle efectivamente a cada clase la oportunidad de aparecer como punto de acceso, y, al mismo tiempo, lograr todas las combinaciones posibles de las diferentes clases incluidas en el encabezamiento.

Observemos ahora el diagrama que se muestra en la figura 20. Aquí vemos cómo en un esquema clasificatorio, o en una lista de encabezamientos, o en un tesoro, el género "Productos manufacturados" puede subdividirse hasta alcanzar gran especificidad. Se destacan tres facetas: material, forma, proceso de fabricación. Cada faceta o aspecto se ha subdividido hasta alcanzar el necesario nivel de detalle.

Es posible llegar a un nivel tan específico como "acero al cromo", "tubo hexagonal", pero no es posible expresar, o recuperar directamente, un documento que trate sobre los "tubos redondos de acero producidos por fundido constante".

Un índice pre-coordinado puede proporcionar diferentes pun

tos de acceso a un documento, tantos como se desee, a partir de los términos de indización asignados a éste. Pero, como las entradas deben seguir una secuencia lineal, estos índices resultan rígidos, inflexibles. Las complejas relaciones entre los temas discutidos en un documento no son unidimensionales, y no es fácil reducirlos a un método lineal de representación.

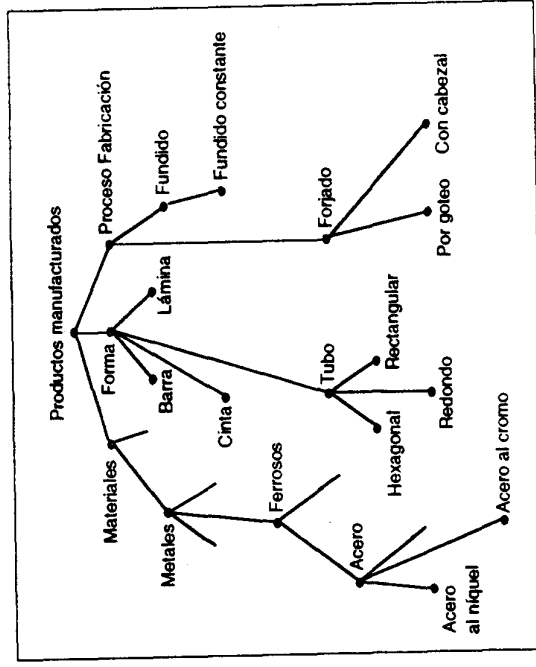


Fig. 20: Ejemplo de subdivisión en un esquema clasificatorio.

6.6. LOS INDICES POST-COORDINADOS.

De acuerdo a lo visto hasta aquí podemos concluir que, para una recuperación eficiente, es necesario contar con un sistema en el que la recuperación del documento sea factible de realizar por cualquier término o clase, y por cualquier combinación posible de ellos. Vale decir que, si existe un documento al que se han asignado los términos A, B, y C, para expresar una determinada conjunción temática, debe ser posible recuperarlo por A, por B o por C, pero también, por AB o AC o BC o ABC.

Ahora, la secuencia lineal ya no tiene significancia en la relación entre las clases, puesto que para estos propósitos, la combinación ABD es idéntica a BDA. En ambos casos están presentes las tres clases requeridas.

En un índice pre-coordinado, como hemos visto, no es posible escapar a las limitaciones de una secuencia lineal. Al construir la entrada en este índice, poniendo juntos los cuatro términos que representan las cuatro clases, se construye una entrada que implica una secuencia lineal de clases: A-B-C-D. (figura 21).

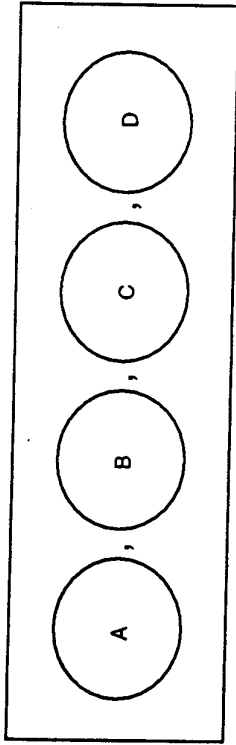


Fig. 21: Secuencia lineal de los términos en un tema complejo.

De hecho, este tipo de entrada puede construirse para representar el más complejo de los temas, pero ésta no es en sí una solución para el problema de la recuperación de información.

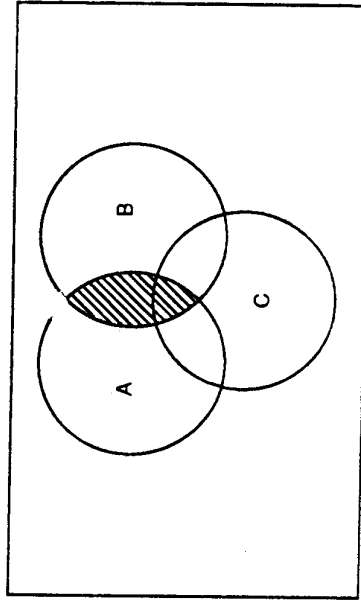


Fig. 22 Diagrama que muestra la intersección de clases

Si observamos la figura 22, vemos que las relaciones **multidimensionales implícitas** en este diagrama, donde la zona sombreada muestra exactamente la intersección de clases, no pueden reducirse a una sola entrada en un índice pre-coordinado.

Por lo tanto, lo que se necesita, para una **recuperación eficiente**, es un sistema en el cual **todas las clases formadas durante el proceso de indización temática del documento, sean igualmente accesibles**, donde no esté involucrada ninguna secuencia de clases, y donde sea fácil confrontar los documentos contenidos en cada clase. En otras palabras, necesitamos un sistema donde sea posible descubrir con facilidad cuáles documentos pertenecen simultáneamente a la clase A y a la clase B. Para satisfacer estos requerimientos se necesita contar con **sistemas de recuperación post-coordinados**.

6.6.1. LOS SISTEMAS PEEK-A-BOO, UNITERMINO Y DE PERFORACION MARGINAL.

Los primeros sistemas verdaderamente **post-coordinados** fueron introducidos en los años 40, por **Batten** en Inglaterra y por **Cordonier** en Francia. El principio en que se basan ha recibido varios nombres, pero el más conocido es sistema "perk-a-boo" o de "coincidencia óptica".

En el **Sistema PEEK-A-BOO**, cada clase determinada en el proceso de indización está representada por una tarjeta, con el término escrito en su borde superior. La superficie restante está dividida en tal forma que cada cuadrado (presumiendo un arreglo tipo matriz), está dedicado a registrar, por medio de perforaciones, el número de un documento o registro en particular, vale decir, hay una posición identificable 7, una posición identificable 25, etc. (figura 23).

Cuando un documento ha sido indizado, las tarjetas que representan las clases a las que fue asignado, por ejemplo, "Medicina Tradicional", "Curanderos" y "Chile", se retiran del archivo y se

perforan en la posición que representa el número identificatorio del documento. Por ejemplo, las perforaciones en las posiciones 12, 94, 182 y 305 en la tarjeta CURANDEROS, indican que todos los documentos identificados por estos números, han sido indizados bajo el término "Curanderos".

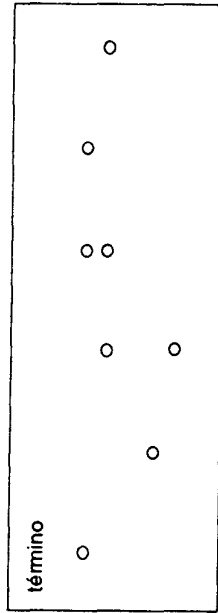


Fig. 23: La Tarjeta peek-a-boo.

El archivo de búsqueda del sistema peek-a-boo consiste en el conjunto de estas tarjetas ordenadas alfabéticamente por el término de indización. Para realizar una búsqueda, luego de establecer la correspondiente estrategia, esto es, decidir qué combinación de clases corresponde más estrechamente al tema que interesa, se retiran del archivo las tarjetas apropiadas. Si nos interesa la "medicina tradicional en Chile", por ejemplo, probablemente sacaremos la tarjeta MEDICINA TRADICIONAL y la tarjeta CHILE. Luego, las tarjetas se superponen, lo que hará posible identificar las posiciones comunes, pues las perforaciones coincidirán y dejarán pasar la luz (de ahí el nombre de "tarjetas de coincidencia óptica). (figura 24).

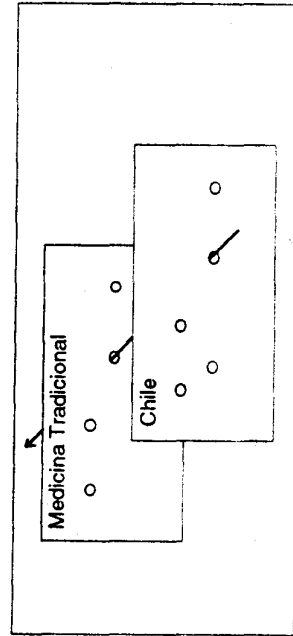


Fig. 24: Búsqueda en el sistema peek-a-boo

Está claro que el principio en que se basa el sistema peek-a-boo satisface los requerimientos que hemos planteado anteriormente para un sistema de recuperación. Se puede buscar por cualquier combinación de clases; es igualmente fácil combinar varias clases que combinar dos. Todas las clases tienen el mismo valor en el sistema; no implica ninguna secuencia de clases. Se puede lograr, en forma simple, el enfoque multidimensional deseado, tanto en la descripción como en la recuperación del tema contenido en el documento, por complejo que éste sea.

El Sistema UNITERMINO, creado por Mortimer Taube en 1951, es también un sistema post-coordinado, basado en la misma idea que el peek-a-boo, pero algo diferente en la forma. Su archivo de búsqueda está formado por tarjetas de 5 por 8 pulgadas, divididas en 10 columnas numeradas del 0 al 9, con un espacio libre en el borde superior por anotar el correspondiente término de indización. Se utiliza una tarjeta por cada término usado y se almacenan ordenadas alfabéticamente. (figura 25).

0	1	2	3	4	5	6	7	8	9

Fig. 25: La tarjeta Unitérmino.

Las columnas verticales se utilizan para anotar los números de los documentos (o de los registros) que han sido asignados al término anotado en la tarjeta. Los números se registran en la columna correspondiente al último dígito.

Para realizar una búsqueda, el procedimiento es similar al del peek-a-boo, pero aquí se comparan visualmente las columnas, buscando aquellos números que coinciden en las correspondientes tarjetas. Los números coincidentes identifican los documentos que tienen en común todas las clases buscadas y, por lo tanto, es dable presumir que tratan sobre el tema de la pregunta o solicitud de búsqueda.

Los sistemas peek-a-boo y unitérmino tienen algo importante en común: ambos operan sobre el principio de una tarjeta por cada clase o término de indización utilizado. A estos sistemas suele llamarse "sistemas documento en el término" (item-on-term-systems), porque la representación de un ítem, vale decir de un documento, se registra EN la tarjeta que representa el término.

También se dice que son "archivos invertidos", porque la relación existente entre los términos y el ítem, ha sido invertida en la organización del archivo, (a diferencia de como aparecen en la hoja de trabajo del indizador o en la tarjeta de registro bibliográfico, donde están registrados los descriptores o términos asignados al documento).

Existe otra forma, completamente diferente, de organizar el archivo. Se trata del "archivo directo" o "término en el documento" (term-on-item). Aquí, cada entrada en el índice representa al documento y en esa misma entrada se listan los términos de indización.

El primer sistema "término en el documento" fue creado por Melvin Mooers en 1948 e implementado en sus "Zator cards", hoy conocidas como **TARJETAS PERFORADAS EN EL BORDE** o "Tarjetas de Perforación Marginal".

La figura 26 ilustra el principio en que se basa este sistema. Cada tarjeta representa un documento. Las perforaciones están cerca de los bordes de la tarjeta. Tales perforaciones, solas o en combinación, son códigos representativos de las clases.

La cita bibliográfica del documento (y, eventualmente, un resumen) están dactilografiados en el frente de la tarjeta.

Las perforaciones que corresponden a los términos de indización asignados al documento, se "abren" por medio de muescas. En el ejemplo que muestra la figura 26, el documento que representa la tarjeta ha sido indizado bajo dos términos: COOPERACION BIBLIOTECARIA Y CHILE, y aparecen marcados los códigos correspondientes.

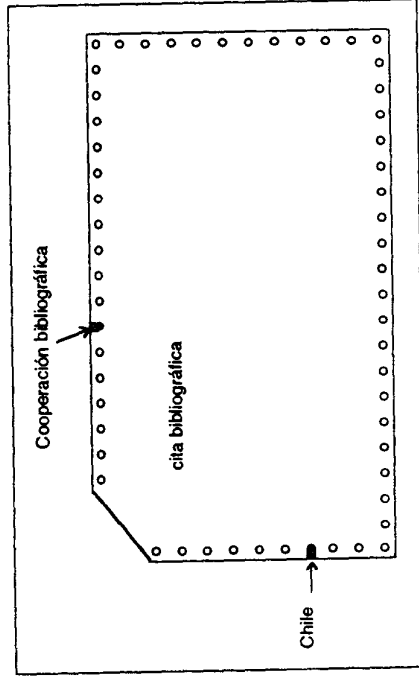


Fig. 26: Tarjeta de borde perforado.

El archivo de búsqueda está constituido por el conjunto total de estas tarjetas. Como éstas no tienen, ni necesitan, ningún orden en particular, al realizar una búsqueda, es necesario revisar el archivo completo.

Supongamos una búsqueda sobre COOPERACION BIBLIOTECARIA Y CHILE. Se inserta una varilla en la posición correspondiente al código "Cooperación Bibliotecaria", haciendo que atraviese el mazo completo de tarjetas. Al levantar la varilla y realizar un movimiento vibratorio, caerán del mazo aquellas tarjetas que tienen una muesca en esa posición. Se toman ahora las tarjetas que cayeron, y se pasa la varilla por la perforación correspondiente al código "Chile". Todas las tarjetas que caigan de este montón han sido "muesqueadas" en ambas posiciones y, por lo tanto, presumiblemente, tratan sobre "Cooperación Bibliotecaria en Chile".

Lo importante de destacar aquí, es que los tres sistemas descritos tienen igual capacidad para combinar las diferentes clases que conforman un tema compuesto. Por otra parte, aunque hoy en día estos sistemas son muy poco usados y casi constituyen historia, el estudio de sus características es útil para comprender mejor la organización de los archivos directos e invertidos con que funcionan las actuales bases de datos computerizadas.

La **matriz documento-término** que se muestra en la figura 27, ilustra claramente las dos formas de organización de archivo que hemos mencionado.

En el diagrama, las letras A-E representan un vocabulario de términos de indización, y los números 1-10 representan una colección de documentos indizados en un sistema de recuperación. La asignación a determinadas clases está indicada por la presencia de una x en la correspondiente celda de la matriz.

	1	2	3	4	5	6	7	8	9	10	documentos
A	x			x					x		x
B									x		x
C	x		x			x		x			x
D									x		
E										x	
términos											

Fig. 27: Matriz Documento-Término

Observando el ejemplo que representa esta matriz, podemos decir que, desde el punto de vista de un archivo directo, el documento N° 10 ha sido asignado a las clases A, B, C y que, desde el punto de vista de un archivo invertido, al término D han sido asignados los documentos 3 y 7.

Por lo tanto, de acuerdo con lo anterior, un archivo de búsqueda es, esencialmente, una matriz documento-término, pues indica qué documentos han sido asignados a qué clase y viceversa.

Las formas de organización de archivos —documento en el término y término en el documento— desarrolladas en los años 40, siguen siendo las formas básicas de organización de archivos utilizadas en los modernos sistemas computerizados de recuperación de información. De hecho, cualquier sistema desarrollado desde la década del 50 en adelante, puede considerarse una **versión automatizada** de los creados por Batten y Mooers. Como hemos dicho, en la actualidad, estos sistemas post-coordinados manuales han sido desplazados por las bases de datos bibliográficos, fácilmente instalables en un microcomputador.

7. EL PROCESO DE INDIZACION

Hemos planteado la indización como un proceso clasificatorio porque nos permite agrupar los documentos relacionados, apartándolos de los no relacionados. Debemos decir también, que la indización es un proceso intelectual mediante el cual se analiza el contenido informativo de un documento (análisis) para luego expresarlo en un lenguaje de indización determinado (síntesis).

Esto concuerda, además, con lo expresado anteriormente, en el sentido que la indización involucra dos pasos básicos: a) decidir de qué trata de un documento, b) traducir tal decisión a términos de indización.

Al igual que el resumen, la indización es un producto del análisis conceptual de un documento, aquél escrito en un lenguaje libre, éste expresado en un vocabulario controlado adoptado por la unidad de información, con el propósito de facilitar la recuperación de los documentos analizados.

7.1. CARACTERISTICAS DEL INDIZADOR.

Si, en un sentido estricto, la indización es un proceso intelectual, entonces, tal proceso debe ser realizado por un ser humano. Pero, para que esta acción sea realizada satisfactoriamente, es natural que esta persona, llamada indizador o analista, necesite reunir ciertas condiciones.

Las características que a continuación se mencionan son algu-

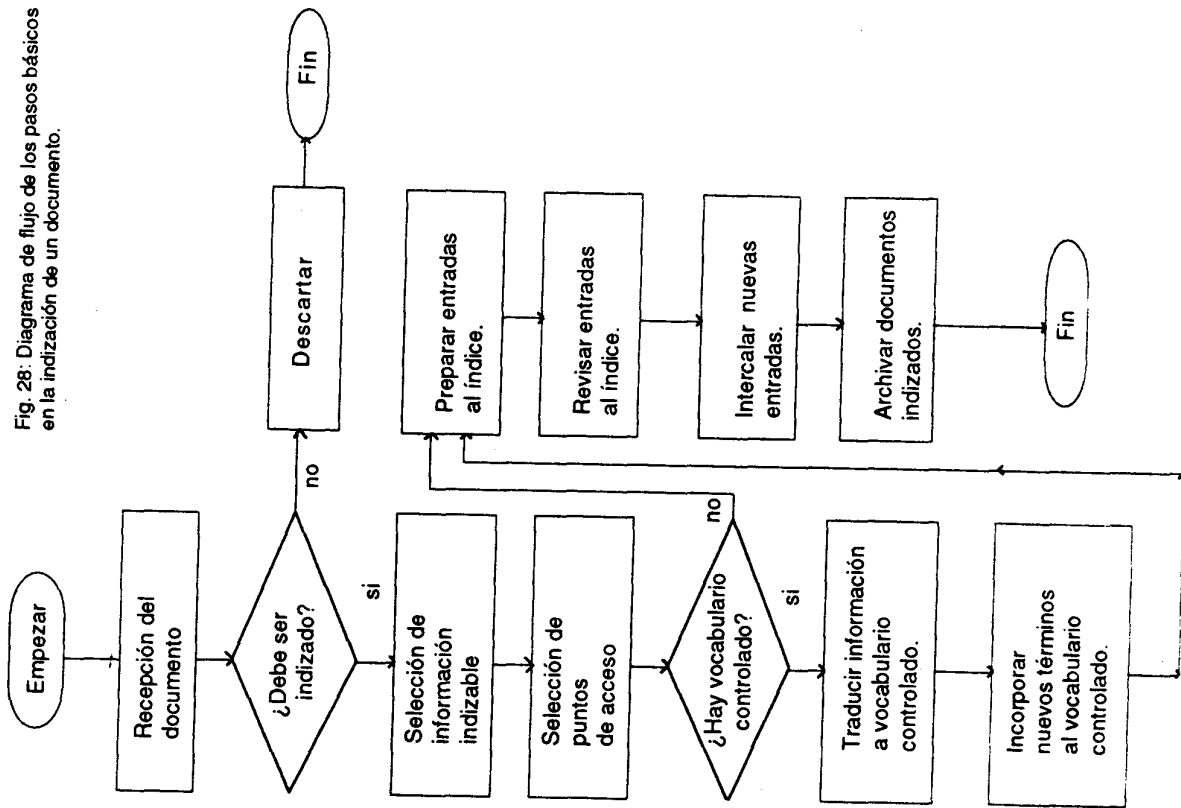
nas de las que la experiencia hace aparecer como las más importantes:

- 1) Conocimiento del tema. A este respecto, algunas personas piensan que la indización sólo puede ser realizada por especialistas en el tema. Otros, en cambio, creen que es suficiente un buen manejo de las técnicas de indización. Ambos extremos parecen exagerados. Sin duda que un conocimiento general del tema y su terminología, más un buen dominio de las técnicas de indización, permiten al indizador realizar un buen trabajo.
- 2) Conocimiento acabado del lenguaje de indización que se utiliza. Evidentemente éste es un aspecto muy importante, e incluye un conocimiento de la terminología, estructura del vocabulario, características, etc.
- 3) Conocimiento de las técnicas y procedimientos de recuperación de información que utiliza el sistema. Considerando que los procesos de indización y recuperación son absolutamente interrelatados, éste es un punto clave para el buen éxito del proceso de indización.
- 4) Conocimiento de los requerimientos y características del usuario. En el fondo esto significa que el indizador debe tener la capacidad de anticiparse a la búsqueda. Implica la necesidad de hacerse preguntas como: ¿Qué aspectos de este documento son de probable importancia para un usuario potencial? ¿Cómo debería yo expresar estos aspectos para lograr una mayor probabilidad de coincidencia entre estas expresiones y las del usuario al momento de la búsqueda?

7.2. PASOS BASICOS EN LA INDIZACION DE UN DOCUMENTO.

El diagrama de flujo de la figura 28 muestra, esquemáticamente, la secuencia de los pasos que suelen darse, en mayor o menor detalle, en la indización de un documento.

Fig. 28: Diagrama de flujo de los pasos básicos en la indización de un documento.



a) **Evaluación del documento.**

Aún cuando no forma parte del proceso de indización propiamente tal, debemos mencionar aquí este paso, que implica la decisión respecto a si un documento debe o no ser ingresado al sistema. Vale decir, es necesario determinar si el tipo de información contenida en ese documento es relevante para la unidad de información, de acuerdo a sus políticas de indización en relación al tema, tipo de documento, fecha de publicación, etc.

b) **Revisión del documento.**

Este paso se refiere a la revisión intelectual o análisis conceptual, con el propósito de extraer ideas o conceptos. En otras palabras, determinar de qué trata el documento. Si previamente se ha preparado un resumen, esta parte del trabajo estaría adelantada en gran medida.

c) **Selección de los puntos de acceso.**

Este paso significa una selección primaria de aquellos términos, extraídos del resumen o del documento, que se consideran como posibles términos de indización, a los que llamaremos "términos candidatos".

d) **Traducción de los términos candidatos al lenguaje de indización.**

Aquí los términos candidatos son traducidos al vocabulario controlado o lenguaje de indización en uso en el sistema. Es importante destacar que, aunque los pasos básicos son los mismos y el mecanismo mental también es el mismo, hay variaciones importantes que dependen de la herramienta de indización que se utilice: es quema de clasificación, lista de encabezamientos de materia, tesauruso; y del tipo de índice que se está construyendo, esto es, si es pre o post-coordinado, si es manual o computarizado.

e) **Incorporación de nuevos términos.**

Cuando las circunstancias lo exigen, puede ser necesaria la creación e introducción de nuevos términos al vocabulario controlado. Esta es una decisión seria que requiere pensamiento y análisis.

sis. Aspectos muy importantes a considerar aquí se refieren a si la unidad trabaja en forma aislada o si pertenece a una red; si es responsable del desarrollo y mantenimiento del vocabulario o si esta tarea es responsabilidad de otra institución.

f) **Intercalar nuevas entradas al índice.**

Naturalmente, este paso adoptará diferentes formas dependiendo del tipo de archivo de búsqueda que se posea. Variará, especialmente, si se trata de un sistema manual como un catálogo en fichas o si se cuenta con una base de datos computarizada.

8. LA INDIZACION COORDINADA.

Como dijimos previamente, al referirnos a los diseñados por Batten, Taube y Mooers, estos sistemas por ser **post-coordinados**, son más flexibles que los catálogos alfabéticos y que los clasificados, porque no dependen de una secuencia lineal. Es más, los sistemas post-coordinados permiten la libre combinación de las clases al momento de la búsqueda. Los pre-coordinados, en cambio, son rígidos porque la combinación de las clases debe hacerse al momento de la indización. (figura 29).

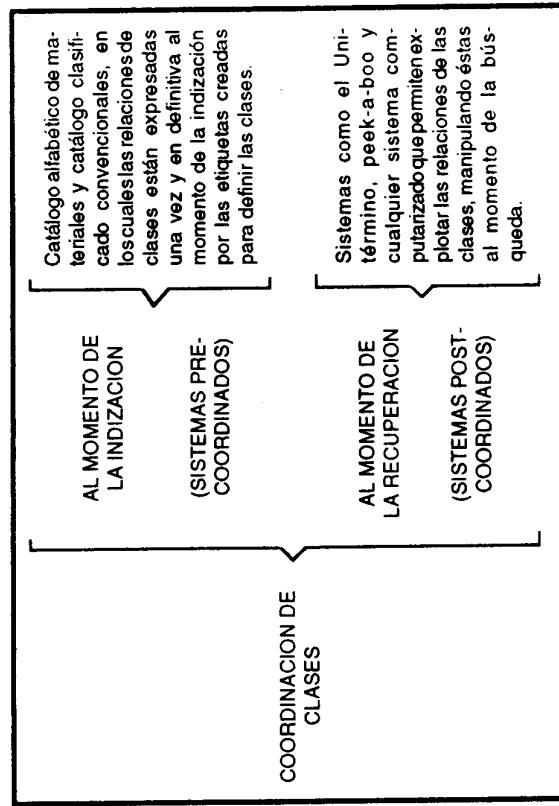


Fig. 29: Sistemas coordinados

Lógicamente, hay diferencias en la indización que se realiza para un sistema u otro, aún cuando el proceso intelectual sea el mismo. En los sistemas post-coordinados la coordinación de los términos se realiza al momento de la búsqueda, pero la inclusión de los términos adecuados para que ello sea posible, se hace al momento de la indización. Vale decir que, siendo la indización y la recuperación procesos mutuamente dependientes, no es posible construir una determinada combinación de clases al momento de la recuperación 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 el **álgebra booleana** o **álgebra de Boole**, (*) que es una forma de álgebra reticular.

El álgebra de Boole, incluyendo sus generalizaciones respecto a la teoría de conjuntos, es una estructura matemática abstracta simple que muestra estructuras de relación, y es justamente este último aspecto el utilizado en relación con la indización.

Sin ir más allá de lo superficial en los aspectos matemáticos del asunto, es posible aplicar las conclusiones de Boole, en ventaja de la indización moderna.

Para ilustrar cómo se producen las relaciones entre los términos, utilizaremos los llamados **Diagramas de Venn**, los que permiten visualizar mejor las operaciones propias del álgebra de Boole.

Previamente a que analicemos las diferentes operaciones booleanas, es necesario decir que el proceso de indización, así como la distinción entre sistemas pre y post-coordinados, son más fáciles de comprender cuando se examinan en el contexto de la teoría de conjuntos.

Un **conjunto** es una colección de objetos llamados "elementos del conjunto".

Hay dos formas de describir un conjunto: a) listando todos los elementos del conjunto: o, b) describiendo una propiedad o característica común a todos los miembros del conjunto.

En la **recuperación de información bibliográfica**, nos encontramos con conjuntos de documentos 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.

En un trabajo de Susan Artandi (1) aparece el siguiente ejemplo para ilustrar la pertenencia simultánea de un documento a varios conjuntos.

Un documento sobre "la terapia en base a penicilina", publicado por J. Smith en 1965, puede ser visto como 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)

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

(*) George Boole (1815-1874): matemático inglés.

Volviendo a los diagramas de Venn, veamos cómo pueden utilizarse para graficar las diferentes **operaciones booleanas**. Supongamos una colección de documentos que llamaremos "M" y que representaremos con un rectángulo, (Figura 30).

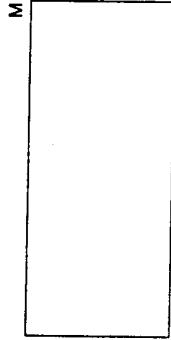


Fig. 30: Diagrama de Venn representando una colección de documentos.

Si revisamos uno por uno todos los documentos en M y a algunos de ellos les asignamos el término de indización "A", hemos creado dos clases:

- 1) La clase que llamaremos A, compuesta por todos los documentos que decidimos indizar con el término A, y
- 2) La clase que llamaremos A', formada por los documentos que decidimos no indizar con el término A.

La clase A' es el **complemento** de A. Y, a su vez, la clase A es el complemento de A'. Es importante recordar aquí, que una clase no puede tener un complemento a menos que se haya tomado una decisión acerca de la **pertenencia o no pertenencia** a dicha clase, de todos y cada uno de los items en M. El siguiente diagrama de Venn ilustra esta formación de clases. (Figura 31).

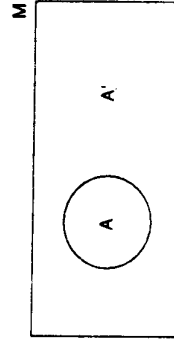


Fig. 31: Diagrama de Venn que ilustra la relación de complementariedad.

La relación entre la clase A y la clase A' es de **complementariedad**. La operación de formar X de X' o X' de X, es lo que se conoce como **complementación booleana**. Se le llama también "Operación Booleana NOT" y se la indica con el siguiente símbolo: $\bar{\square}$

Podemos ejemplificar la relación de clases así formadas de la siguiente manera: Supongamos que la clase A ha sido formada asignando el término PSICOLOGIA a un número de documentos en la colección. Por cierto no podemos decir, estrictamente, que son documentos que tratan sobre "psicología". Simplemente, es una clase formada por aquellos documentos en la colección a los que se decidió asignar el término PSICOLOGIA. Naturalmente, es lógico esperar una alta y marcada coincidencia entre los documentos en la colección que efectivamente tratan sobre psicología y aquellos incluidos en la clase PSICOLOGIA.

Lo importante, entonces, es que la clase complementaria A' es la clase formada por los documentos que, **después de examinados**, no fueron asignados al término PSICOLOGIA.

Supongamos, ahora, que formamos otra clase, a la que llamaremos B, asignando el término de indización "B" a una cantidad de documentos y no asignándolos al resto. (figura 32). Esto crea otras dos clases:

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

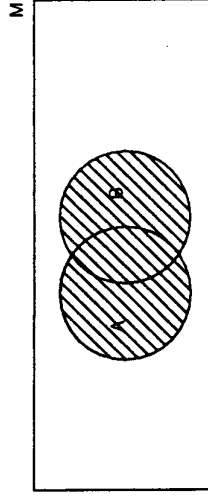


Fig. 32: Suma Lógica.

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

El reconocimiento de esta suma de dos o más clases nos permite posteriormente, al momento de la búsqueda, pedir documentos que hayan sido indizados por uno u otro término indistintamente. Por ejemplo, (PSICOLOGIA \cup PSICOANALISIS \cup PSICOTERA-PIA).

Se ha formado también una clase de documentos que pertenecen a **ambas** clases, A y B, simultáneamente. (figura 33). Hay una pertenencia común o conjunta del documento, a A y a B, conocida como "intersección", "producto lógico", "operación booleana AND", y que se expresa con el símbolo: \cap

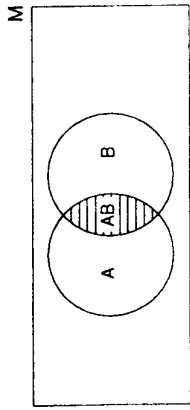


Fig. 33: Producto Lógico.

El reconocimiento de esta intersección de clases nos permite posteriormente, al momento de la búsqueda, solicitar documentos cuya característica sea pertenecer **simultáneamente** a más de una clase.

Por ejemplo, podemos pedir documentos que pertenezcan a la clase **TRANSPLANTE** (los habrá de muchos tipos) Y a la clase **RIÑON** (habrá documentos sobre diferentes problemas del riñón); vale decir, documentos que hayan sido asignados a ambas clases, indizados bajo ambos términos, y que, presumiblemente, traten sobre "transplante renal" o "transplante de riñón" (fig. 34).

En este punto es aconsejable recordar que todas las relaciones de clases que hemos descrito, tienen validez y significado **si y sólo si** hemos tenido en cuenta las condiciones antes dichas:

(a) definitivamente, a todos y cada uno de los items en la colección se ha asignado, o no asignado, todos los términos de indiza-

ción que correspondan;

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

Si no se cumple la primera condición, las operaciones que se realicen con las clases serán ambiguas o sin significado. Si se olvida la segunda, las operaciones, aunque válidas, puede que no tengan nada que ver con el significado que les dé el usuario.

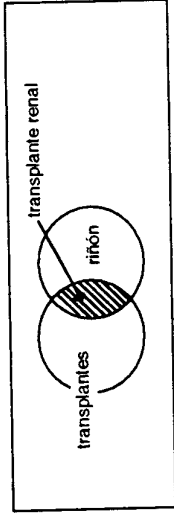


Fig. 34: Intersección de clases.

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 $\bar{\cap}$

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

Existe otra relación de clases muy importante que es la "inclusión de clases". Esta ocurre si todos los miembros de una clase son a su vez miembros de otra. Por ejemplo, la clase **CHILENOS** está incluida en otra clase más grande que es **LATINOAMERICANOS**. (figura 35).

La principal relación de clases que aparece en una clasificación jerárquica es la de **inclusión**. De hecho, cualquier par de clases pueden ser **disyuntivas** (no tienen miembros en común) o **inclusivas** (una incluye a la otra). (figura 36).

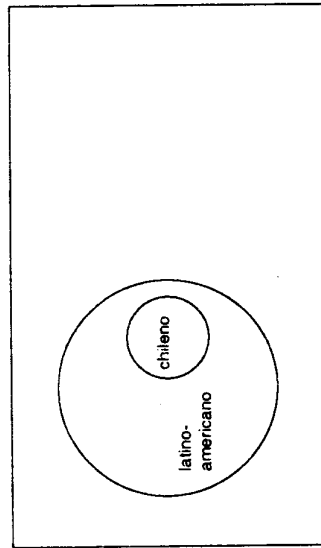


Fig. 35: Relación de inclusión.

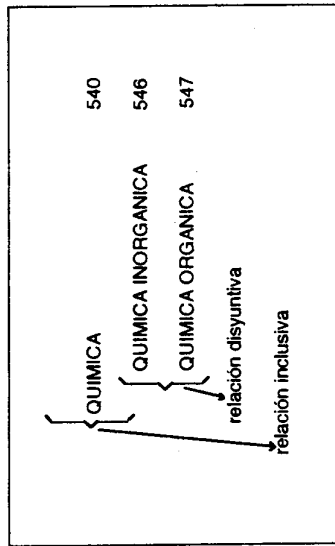


Fig. 36: Relación inclusiva y disyuntiva.

Así, un esquema de clasificación (explícita o implícitamente) consiste en un conjunto de **cadena**s, descendiendo desde un término más genérico a un término más específico.

Al hablar de los índices pre-coordinados en el capítulo 5, en la figura 20 se muestran varias cadenas que van de lo genérico a lo específico. Ahora, en la figura 37 podemos observar una cadena que va desde METALES a ACERO AL CROMO. A cualquier nivel pueden crearse nuevas cadenas de modo que haya "ramificación".

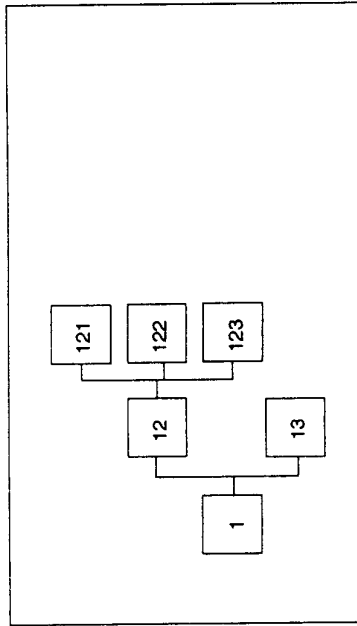


Fig. 37: Cadenas y ramificaciones por áreas.

En la figura 38 aparece la misma cadena pero representada en forma de árbol.

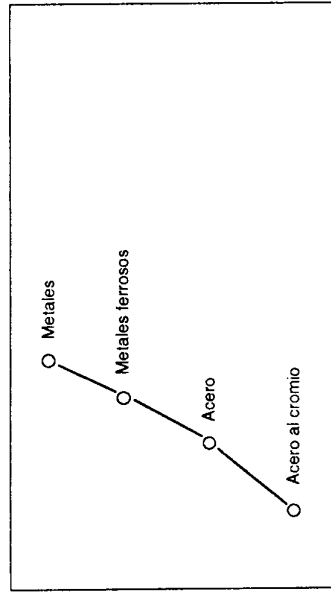


Fig. 38: Cadenas en forma de árbol.

En resumen, podemos decir que los sistemas post-coordinados permiten la coordinación de clases al momento de la búsqueda. Esto nos permite pedir documentos que hayan sido asignados a la clase A pero no a la B (complementación); documentos que pertenecan 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).

9. EXHAUSTIVIDAD VS. ESPECIFICIDAD.

Como hemos dicho, reiteradamente, la labor del indizador es **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.

Así, por ejemplo, la selección podría guiarse por una cierta categorización reconocida como importante en el área temática cubierta por el documento. Pueden ser fenómenos, procesos, propiedades, equipos en el área tecnológica, o determinadas variables en el campo de las ciencias sociales.

Supongamos un documento que trata sobre "uso de fertilizantes en la agricultura". Es probable que el indizador busque **sistemáticamente**, conceptos como "tipos de cultivo para los que sirve", "efectos secundarios negativos", "formas de aplicación del fertilizante", etc.

Los aspectos mencionados suelen afectar a dos **características** de la indización, las que a su vez afectan directamente la posterior recuperación de información. Estas características de la indización son: la exhaustividad y la especificidad.

9.1. EXHAUSTIVIDAD.

Se llama **exhaustividad en la indización** a la medida de la extensión en que 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.

Veamos el siguiente ejemplo. Al analizar un documento describimos que trata sobre una diversidad de temas complejos, "Análisis histórico en 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. Plantea, también, que de estos últimos ninguno fue a bibliotecas escolares, prefiriendo las bibliotecas universitarias".

Supongamos que se reconocen los siguientes temas, los que, luego de traducidos al vocabulario del sistema, a su vez se transforman en términos de indización.

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

Se puede decir que la indización de este documento ha sido altamente exhaustiva. Obviamente, si todos estos temas son indizados, al momento de la búsqueda el documento puede ser recuperado por cualquiera de estos temas o clases o combinación de ellos.

Este hecho (esta recuperabilidad del documento) nos permite decir que un **alto nivel de exhaustividad, asegura un alto nivel de "recall" o recuperación.** (*)

(*) Utilizamos la palabra "recall" en la expresión inglesa, debido a que no hemos encontrado consenso en su traducción al español. Algunos autores le llaman "utilidad", otros simplemente "recuperación". A fin de evitar ambigüedades, hemos preferido privilegiar la claridad del concepto por sobre la pureza del español.

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

Sin embargo, el alto nivel de exhaustividad, si bien favorece el "recall", tiende a afectar negativamente, a disminuir la precisión en la recuperación, al 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.

La exhaustividad afecta negativamente a la precisión en la recuperación, básicamente por dos razones:

(a) Si para cada documento ingresado reconocemos todos o casi todos los diferentes temas tratados, ese documento será recuperado por cualquiera de esos puntos de acceso (favoreciendo el recall). Es probable, sin embargo, que se estén incluyendo temas en la indización que, si bien aparecen en el documento, estén tratados en forma superficial o tangencial. Pero, igualmente el documento será recuperado en respuesta a preguntas sobre cualquiera de estos temas o combinación de ellos.

(b) La otra razón por la cual la exhaustividad disminuye la precisión, es que 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.

El documento del ejemplo ha sido indizado bajo 11 términos diferentes. Esto significa que se puede recuperar por cualquiera de estos términos o combinación de ellos; pero, también, puede ser erróneamente recuperado por una falsa coordinación entre los términos.

Los siguientes ejemplos de preguntas muestran cómo este documento, al ser recuperado, puede no responder a las expectativas

del usuario, ya sea porque los temas son tratados muy tangencialmente, o porque se ha producido una falsa coordinación entre los términos.

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

En estos cuatro casos la recuperación de este documento en particular ha resultado inadecuada para las necesidades del usuario.

En resumen, podemos mencionar 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 cada unidad, pero también riesgosa.

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

9.2. ESPECIFICIDAD.

A diferencia de la exhaustividad, que es una decisión propia de una política de indización y que afecta positivamente la capacidad de recall del sistema, la **precisión** de una búsqueda, en cambio, depende en gran medida de la **especificidad del lenguaje de indización**. Esta especificidad radica en la capacidad del lenguaje para describir en forma específica, precisa, los diferentes temas tratados en los documentos.

En tres unidades de información diferentes, con tres vocabularios con diferente nivel de especificidad (en relación al tema), puede suceder el siguiente fenómeno con un mismo documento que trata, por ejemplo, sobre **"Secuelas de la difteria"**.

- En la Biblioteca A (con un vocabulario bien específico, de 2000 términos o más) el documento podría ser indizado bajo: DIFTERIA-SECUELAS, o SECUELAS DE DIFTERIA, o SECUELAS Y DIFTERIA.
- En la Biblioteca B (regular especificidad, vocabulario de 1000 términos), podría indizarse bajo: ENFERMEDADES TRANSMISIBLES.
- En la Biblioteca C (vocabulario muy general, no más de 500 términos) tal vez estaría indizado bajo ENFERMEDADES.

En el siguiente diagrama podemos observar el efecto que estas variaciones en la especificidad del vocabulario tendrán sobre la recuperación. (figura 39)

Al momento de la búsqueda, la situación será diferente en las tres bibliotecas:

En la Biblioteca A, con vocabulario muy específico, es posible expresar exactamente el tema de interés: "Secuelas de la difteria", por lo que el conjunto de documentos recuperados tiene una alta probabilidad de ser relevante, logrando un alto **nivel de precisión**. Sin embargo, la relación de recall será más baja, pues es posible que hayan quedado en la colección sin recuperar documentos que

traten sobre "las secuelas de la difteria", pero que fueron indizados bajo "enfermedades transmisibles", por ejemplo. Y así sucesivamente, según se observa en el diagrama.

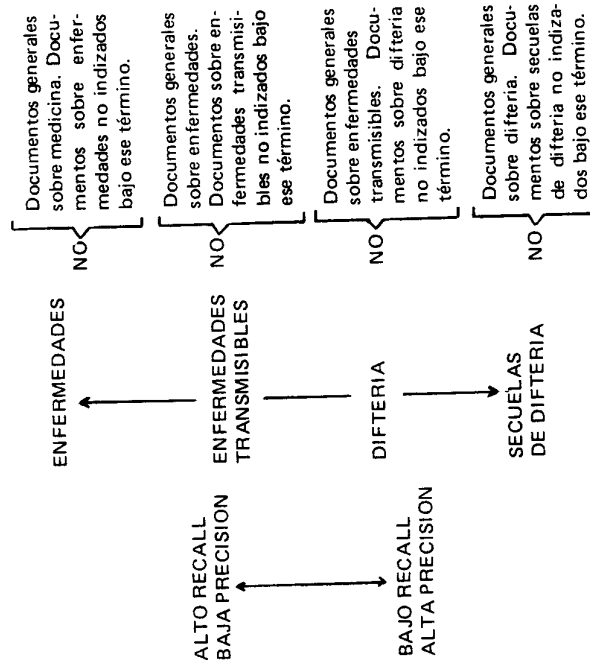


Fig. 39: Relación inversa recall/precisión.

En **resumen**, la exhaustividad en la indización controla la capacidad de "recall" de un índice, mientras que la especificidad de un lenguaje de indización incide directamente en la capacidad de precisión de un índice.

ducción a términos de indización. En otras palabras, primero es preciso decidir de qué trata un documento o pregunta y luego se procura traducir tal análisis conceptual a los términos del lenguaje de indización.

La etapa de análisis y la etapa de traducción tienen diferente efecto sobre el performance del sistema, como se vió en el capítulo 9 al hablar de la exhaustividad y la especificidad. Por ejemplo, después de un acucioso análisis, el indizador decide que un documento trata sobre "evaluación de proyectos de información". Este es el resultado de su análisis conceptual. Pero, cuando intenta traducir este análisis al lenguaje de indización, puede que no encuentre los términos precisos y tenga que describir el tema en una forma más general, con términos menos específicos. El vocabulario del sistema no le permite identificar exactamente el tema "evaluación de proyectos de información", de tal manera que, tal vez, deba indizar el documento bajo "Evaluación de Proyectos", o simplemente bajo "Proyectos".

Debido a la aparente simplicidad del sistema Unitérmino, Taube y sus tarjetas tuvieron gran éxito. Como no se enfatizaba la necesidad de un vocabulario controlado, se creó una suerte de mito: el indizar documentos para el sistema Unitérmino era una tarea rutinaria y simple que podía realizar cualquier persona con un poco de entrenamiento, la que se guiaba por la presencia de determinadas palabras en el texto. No había control de sinónimos, ni homónimos, ni existían términos compuestos.

Además de los problemas bastante obvios que acarrea la falta de control de homónimos y sinónimos, el utilizar un vocabulario sin pre-coordinación alguna, por lo general, conduce a una frecuente "falsa recuperación". Por ejemplo, si se necesitan documentos sobre "administración de bibliotecas universitarias", en el unitérmino primitivo, habría que buscar por los siguientes términos, para realizar la consiguiente post-coordinación:

ADMINISTRACION Y
BIBLIOTECAS Y
UNIVERSIDAD

10. EL CONTROL DEL VOCABULARIO

Generalmente, se conoce como **vocabulario controlado** o **lista de autoridad**, a una lista autorizada de términos de indización que se utiliza en una unidad de información para propósitos de la indización temática de los documentos.

Lancaster (1) plantea que la utilización de vocabularios controlados, tiende a mejorar la consistencia de la indización. Es más probable que dos indizadores coincidan en los términos elegidos para describir un tema, si esa elección la hacen a partir de una lista pre-establecida, a que si ellos, independientemente, crean los términos según los necesiten. Además, como ya se ha expresado en más de una oportunidad, el rol de este vocabulario es extremadamente importante para poder crear perfiles de documentos y de preguntas que sean confrontables entre sí.

Otro aspecto destacable del uso de vocabularios controlados y bien estructurados, es que facilitan las búsquedas genéricas, ya que estos vocabularios de alguna manera deben indicar la relación jerárquica entre los términos.

El preeminente papel del vocabulario controlado está presente tanto en las operaciones de indización como en las de búsqueda. Pero, es necesario enfatizar que el vocabulario no influye, o por lo menos no debe influir, en el análisis conceptual de los documentos ni de las preguntas. Como vimos en el capítulo 7, la etapa de análisis conceptual del documento está separada de la etapa de tra-

(1) LANCASTER, F.W. *Vocabulary Control for Information Retrieval*. Washington, D.C., Information Resources Press, 1972.

Con esta estrategia es probable que se recuperasen documentos tales como:

D1: Literatura sobre administración en bibliotecas universitarias

(Indizado bajo: LITERATURA, BIBLIOTECAS, UNIVERSIDAD, ADMINISTRACION)

D2: Administración de bibliotecas universitarias

(Indizado bajo: ADMINISTRACION, BIBLIOTECAS, UNIVERSIDAD).

De estos dos documentos sólo uno es relevante a la búsqueda. Es obvio que se produjo una falsa recuperación con el documento N° 2.

Por lo tanto, la forma primitiva del unitérmino duró relativamente poco. El primer cambio fue aceptar términos compuestos, como por ejemplo "corriente alterna", y luego, términos pre-coordinados como "ingeniería química", "recursos naturales".

La utilización de términos compuestos, de más de una palabra, significó que los términos de indización se transformaron en etiquetas usadas para expresar conceptos unitarios, sin importar el número de palabras necesarias para expresarlos correctamente. (figura 40)

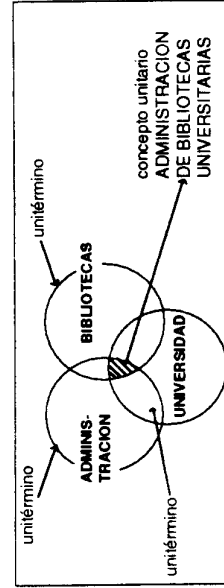


Fig. 40: Etiqueta para un concepto unitario.

Aún cuando conceptualmente, hay una clara distinción entre **sistemas pre-coordinados** y **sistemas post-coordinados**, y es fácil apreciar tal distinción, la situación es menos clara en el plano del lenguaje.

En español, por ejemplo, hay palabras que en sí encierran una pre-coordinación de conceptos o de clases:

TERMOMETRO = Instrumento para medir temperatura

HIDROFOBIA = horror al agua

Pero, esto no ocurre frecuentemente y, en general, los conceptos necesitan más de una palabra para ser expresados.

Actualmente, la mayoría de los sistemas son post-coordinados, (en el sentido de poder post-coordinar las clases al momento de la búsqueda) pero, usan vocabularios con un alto grado de pre-coordinación en sus términos (en el sentido de conceptos unitarios expresados con dos o más palabras). (figura 41).

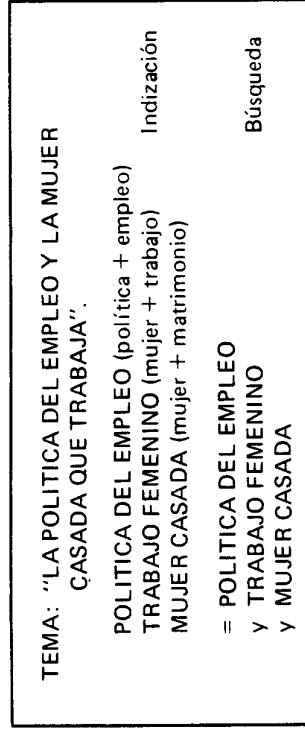


Fig. 41: Ejemplo de sistema post-coordinado con vocabulario pre-coordinado.

Un adecuado nivel de pre-coordinación en los vocabularios facilita la indización y la búsqueda en los sistemas post-coordinados y reduce los problemas de falsas o ambiguas combinaciones entre los términos.

Sin embargo, la existencia de otros problemas propios del lenguaje natural, hacen necesaria la existencia de un **vocabulario con-**

trolado con las siguientes características:

- 1.- **Control de sinónimos**, para evitar que documentos similares se indiquen en forma distinta.
- 2.- **Control de homógrafos**, para evitar la confusión entre palabras que se escriben igual pero que tienen distinto significado.
- 3.- **Genéricamente estructurado**, a fin de que muestre las relaciones jerárquicas entre los términos, lo que facilita, al momento de la indización, encontrar los términos más generales o más específicos según se necesite y, además, permite la búsqueda genérica al momento de la recuperación.
- 4.- Con un adecuado **nivel de pre-coordinación** en sus términos.
- 5.- **Notas de alcance**, esto significa la inclusión de notas aclaratorias respecto a algunos términos, las que explican el contexto en que ha sido utilizado el término en ese vocabulario en particular.

Hoy en día, con el uso masivo de los sistemas computarizados, muchas personas piensan que los lenguajes no controlados o la búsqueda en texto libre son altamente deseables para los sistemas de recuperación de información, porque, por una parte, se ahorra tiempo y se simplifica el ingreso de información y, por otra parte, se protegen las expresiones del autor y es posible revelar diferencias culturales, psicológicas o lingüísticas en el texto de un documento.

Pero, también, hay muchos argumentos en pro de la indización utilizando un vocabulario controlado. Debido a la riqueza de las expresiones, los sistemas con lenguaje libre se tornan confusos cuando se los consulta en busca de **conceptos** más que de palabras. Es necesario adivinar todas las posibles formas en que un mismo concepto puede haber sido expresado, y el usuario no siempre gana en este juego de adivinanzas, resultando la búsqueda un proceso largo y complejo.

La tendencia actual más predominante parece ser un sistema mixto, vale decir, la indización de los documentos con un vocabulario controlado y el almacenamiento de esta indización junto a un resumen de buena calidad. Esto permite una complementación al momento de la búsqueda, pues casi todas las bases de datos permiten buscar por palabras claves y también en texto libre.

Como dice Raya Fidel, "la disponibilidad de bases de datos bibliográficos para búsquedas en línea, cambió el rol de los resúmenes. Hoy en día los resúmenes de las bases de datos disponibles al público pueden examinarse con la modalidad de texto-libre, lo cual permite al usuario realizar búsquedas en línea por cualquier término que considere apropiado" (1)

Hemos dicho que los lenguajes de indización son en cierta medida "artificiales". Un lenguaje natural es el lenguaje que la gente "naturalmente" habla. El lenguaje de indización, en cambio, ha sido "inventado" para una forma especial de comunicación, pues intenta que autores, indizadores y usuarios puedan comunicarse entre sí.

La selección o creación del lenguaje de indización, probablemente, es una de las tareas más difíciles en el diseño de un sistema de recuperación de información. Es una tarea que requiere conocimientos lingüísticos, excelente manejo del idioma, acabado conocimiento de las diferentes estructuras y formas que pueden adoptar estos vocabularios y, naturalmente, un profundo conocimiento del área temática a cubrir por el vocabulario. El reconocimiento de las características y requerimientos especiales de esta tarea, evidencia la necesidad de un equipo interdisciplinario para su realización.

(1) FIDEL, Raya. "Writing abstracts for free-text searching". Journal of Documentation. 42(1): 11-21, March 1986.

10.1. REQUISITOS DE UN LENGUAJE DE INDIZACION.

Cualquier lenguaje en que se intente expresar un índice, requiere de algunos atributos básicos:

- a) Debe contar con un **vocabulario** y una **sintaxis** comprensibles para el que busca. Si los usuarios de la biblioteca son público general, o al menos un grupo no especialmente entrenado, el lenguaje de indización debe aproximarse en lo posible al lenguaje natural, o ser fácil de aprender. En una unidad de información extremadamente especializada, el vocabulario debe reflejar las expresiones que son propias de los especialistas.
- b) Algunos archivos de búsqueda se construyen para la búsqueda por computador. Esto significa que el formato, la sintaxis y el vocabulario del lenguaje de indización, deben ser comprensibles para el computador y para el hombre.
- c) Un lenguaje de indización bien diseñado debe ser capaz de aceptar sus propias modificaciones, en la medida que cambia el lenguaje de los documentos o de los usuarios.

10.2. ESTRUCTURA DE LOS LENGUAJES DE INDIZACION.

Principalmente, nos preocupan dos componentes básicos de un lenguaje de indización: el vocabulario y la sintaxis.

Un **vocabulario** es el conjunto de palabras que se usan en un lenguaje. Esto que es bastante evidente en un lenguaje natural, es aplicable a un lenguaje de indización. Aquí el vocabulario es el conjunto de palabras que se usan, solas o en combinación, para construir términos de indización que serán utilizados en la descripción temática de un documento.

La **sintaxis** de un lenguaje es el conjunto de reglas para combi-

nar los elementos del vocabulario en **unidades idiomáticas** con significado no expresable con el vocabulario básico, llamadas **unidades sintácticas**. Si admitimos como sustantivos "árbol" y "agua", y como adjetivos "verde" y "azul", usamos la sintaxis para expresar las imágenes de "árbol verde" y "agua azul" que no puede expresarse el vocabulario básico.

Un término de indización, entonces, es una **unidad sintáctica** compuesta de una o varias palabras de alguna forma ligadas entre sí.

En los lenguajes artificiales de indización, una parte muy importante es una **autoridad** para el vocabulario y la sintaxis, ya que la utilidad de estos lenguajes se basa en lo restrictivo de su uso y en que quienes lo usan aceptan sus definiciones formales y sus reglas sintácticas.

Las **listas de autoridad** especifican la pertinencia, estructura, uso, relaciones y significado de los elementos del vocabulario. Incluiremos aquí a los tesauros, esquemas de clasificación, listas de encabezamientos de materia, listados de palabras claves, etc.

La figura 42 muestra una página extraída de una versión moderna y abreviada del tesauro de Roget. (1).

Indice	Texto
SOLILOQUY 589	87. UNITY. --N. unity, oneness; individuality, unification, etc. 48; completeness, completion.
solitary alone 87	one, unit, individual.
solitude 893	V. isolate, insulate.
solo 415	render one: unite, etc. (join),
soluble 335	43, (combine), 48.
...	Adj. one, sole, lone, single, solitary;
...	individual, apart. ...
	589. SOLILOQUY. --N. soliloquy, monologue, apostrophe.
	V. soliloquize, monologize, talk to oneself; think aloud, apostrophize.

Fig. 42: Transcripción de trozos del índice y texto del Tesauro de Roget.

Según Vickery (2), dentro del contexto de la recuperación de información, Helen Brownson fue la primera persona que usó por escrito la palabra "tesauro" en su común aplicación y sentido que hoy le damos. Fue en 1957, en la Conferencia sobre Clasificación realizada en Dorking, Inglaterra, donde ella dijo: "el problema (de la recuperación de información), al menos como algunos investigadores lo ven, es transformar los conceptos y sus relaciones como aparecen expresadas en el documento, a otro lenguaje más reglamentado, con los sinónimos controlados y las estructuras sintácticas simplificadas. . . Es razonable pensar que mientras más avanzamos en la rutinización y mecanización de las técnicas de traducir el lenguaje ordinario a uno reglamentado y codificado para una mani-

(1) ROGET'S Pocket Thesaurus, New York, Pocket Books, 1946.

(2) VICKERY, B.C. "Thesaurus - a new word in documentation" Journal of Documentation, 16(4): 181-189, 1960.

11. LOS TESAURUS.

11.1. INTRODUCCION.

La palabra **tesauro** viene del griego y significa "almacén" o "tesoro". Según Alan Gilchrist (1), el Shorten Oxford English Dictionary registra 1736 como la fecha más antigua conocida en que se usó esta palabra (thesaurus), significando un "tesoro o almacén del conocimiento, como un diccionario o una enciclopedia". El Diccionario Webster va más allá y define al tesauro como "un libro de palabras o de información acerca de un campo en particular, o un conjunto de conceptos, específicamente un diccionario de sinónimos".

Sin duda el tesauro más famoso ha sido el diseñado por Peter Mark Roget en 1852, con el título "Thesaurus of English Words and Phrases". El propio Roget lo define así: "Una colección. . . arreglada, no en orden alfabético como un diccionario, sino de acuerdo con las ideas que expresan. El objeto es, ante una idea dada, encontrar la o las palabras por medio de las cuales esa idea puede ser expresada más precisa y adecuadamente".

Dicho en otra forma, una palabra que no se desea, pero que se aproxima a la idea que se quiere expresar, se busca en un índice alfabético que refiere al lector hacia uno o más grupos de palabras, conceptualmente relacionadas, entre las cuales puede encontrar una palabra que prefiera porque expresa mejor su idea. Como vemos, el tesauro de Roget es más bien un diccionario de ideas afines.

(1) GILCHRIST, Alan. The Thesaurus in Retrieval London, Aslib, 1971.

pulación mecánica, más fácil nos será lograr una búsqueda mecánica, económicamente factible, a gran escala. . . Algunos investigadores piensan que la mejor respuesta o solución. . . puede ser la aplicación de un **tesauro** mecanizado basado en redes o significados relacionados”.

Sin embargo, como dice Vickery, probablemente la primera persona en pensar en términos de un tesauro para la recuperación de información fue H.P. Luhn en 1957, quien sugirió, con propósitos de indización, la compilación de “familias y nociones” y “diccionarios de familias nocionales”, muy similar al concepto de Roget.

11.2. DEFINICIONES.

En primer lugar, un tesauro es una forma de vocabulario controlado, al igual que lo es una lista de encabezamientos de materia, pero con características estructurales y de uso muy propias.

Un tesauro se define como: una lista de autoridad o vocabulario controlado de la terminología de algún área específica del conocimiento, consistente en una enumeración de términos de indización aprobados, generalmente llamados **descriptores**, que muestren éstas jerárquicas o de otra índole, así como las necesarias notas de alcance. Incluye también, entradas para términos no aprobados (no-descriptores) en cuyo caso hay una referencia que envía al término aprobado.

Los tesauros combinan características de los esquemas de clasificación y de las listas de encabezamientos de materia. Los términos suelen aparecer ordenados alfabéticamente, pero, al igual que un esquema de clasificación, el tesauro muestra explícitamente las relaciones jerárquicas entre ellos. Por otra parte, mientras los esquemas de clasificación tienden a destacar el conjunto de las relaciones jerárquicas, el tesauro entrega las relaciones necesarias para la indización y la recuperación de información, según el contexto

de la temática que cubre, y las necesidades de los usuarios. (figura 43)

A	312	Demografía
B	312.1	Natalidad
C	312.12	Tasa de natalidad
D	312.121	Tasa específica de natalidad

Fig. 43: Estructuras jerárquicas.

De acuerdo a la Norma Chilena del año 1976 relativa a la construcción de tesauros (1), basada a su vez en otras normas internacionales, los tesauros pueden definirse según su función y según su estructura. Dicha Norma dice, textualmente:

“Desde el punto de vista de su función, un tesauro es una herramienta de control terminológico utilizada para traducir a un lenguaje más restringido (lenguaje documentario, lenguaje de indización), el lenguaje natural usado en los documentos y utilizado por los indizadores y por los usuarios. Desde el punto de vista de su estructura, el tesauro es un vocabulario controlado y dinámico de términos que guardan entre sí relaciones semánticas y genéricas, que se aplica a un campo específico del conocimiento”.

Como un subconjunto estructurado, el tesauro sirve para describir el **contenido temático** de documentos, objetos, obras de arte, o colecciones de datos. Los términos necesarios para describir algunas características formales y otros aspectos del documento u objeto —términos no temáticos tales como nombres de personas, instituciones, lugares geográficos— no necesitan formar parte del tesauro. Estos términos, usualmente llamados **identificadores**, suelen ser necesarios para indizar correctamente un ítem y, por lo tanto, pueden constituir un apéndice del tesauro.

(1) NORMA CHILENA NCh1278. c78. Estructuración y desarrollo de tesauros monolingües destinados a la recuperación de información”. Santiago, INN, 1978.

Un tesoro especializado debe ser capaz de reflejar con precisión la información contenida en el conjunto de documentos u otros ítems que conforman la colección a la cual se le intenta aplicar. Debe contener los términos y referencias apropiados al tema, tomando en consideración tanto el lenguaje de la colección de documentos, como las características y necesidades de información de los usuarios.

En base a la naturaleza del **control terminológico**, se desarrollan principalmente dos tipos de tesauros:

- a) Los tesauros cuyas reglas de control terminológico se basan en la selección de **términos preferenciales**, es decir, tesauros en los cuales se admite un solo término para representar un concepto con miras a la indización y a la recuperación de información.
- b) Los tesauros cuyas reglas aceptan para la indización y la recuperación, **todos los términos** que representan un concepto, haciendo, a su vez, corresponder todos estos términos a una **representación inequívoca** del concepto, ya sea un número, notación, o símbolo. En este tipo de tesauros es más fácil suprimir los sinónimos y modificar las relaciones anteriores cuando un término cambia de sentido.

Si bien los tesauros de elección preferencial pueden actualizarse manualmente, los que aceptan todos los términos exigen el uso de un computador para ser puestos al día.

11.3. ESTRUCTURA DE LOS TESAUROS.

La configuración interna de cada uno de los términos de entrada y la ordenación de estas diferentes entradas en relación mutua, constituyen la estructura de un tesoro.

En un tesoro las referencias cruzadas **explicitan** la naturaleza de las relaciones que ligan mutuamente a las entradas dentro de una red de conceptos.

11.3.1. DESCRIPTORES.

En un tesoro, los términos establecidos para la indización reciben el nombre de **descriptores**. Un **descriptor** puede definirse como un **término o un símbolo autorizado y formalizado que figura en un tesoro y que se usa para representar, sin ambigüedad, los conceptos contenidos en los documentos y en las preguntas**. Pueden considerarse como **descriptores**:

- a) Los términos que representan conceptos o combinaciones de conceptos (descriptores propiamente tales).
- b) Los términos que representan entidades individuales, como lugares geográficos, entidades, siglas, nombres de proyectos, nomenclaturas, marcas de fábrica, etc., usualmente llamados **identificadores** para diferenciarlos de los descriptores propiamente tales.

Pensando en el **intercambio** o cooperación a nivel internacional, se puede distinguir dos niveles de descriptores:

- Descriptores más generales, que pueden ser objeto de acuerdos multinacionales después de ser traducidos a diferentes idiomas;
- Descriptores más específicos que podrían ser objeto de acuerdos en áreas o campos del conocimiento.

Se habla de **microtesoro**, cuando se trata de un tesoro sobre un tema específico que, a su vez, está incluido en un tesoro que cubre un campo más amplio. Vale decir, es la profundización de un tema específico dentro del marco del tesoro mayor.

El concepto de **macrotesoro** es algo más complejo. No sólo implica que su terminología abarca un área amplia del conocimiento, sino que, además, debe permitir que los descriptores sean comprendidos y usados en la misma forma por los miembros de una red, por ejemplo. Además, debe estar estructurado de tal forma, que permita el nacimiento o extensión de otros tesauros a partir de él.

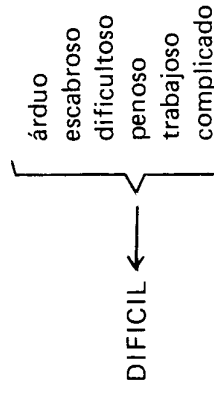
11.3.2. RELACIONES ENTRE LOS TERMINOS.

Una de las características de estas listas de autoridad, llamadas **tesauros**, es que muestran explícitamente diversas relaciones entre conceptos, mediante la indicación de las interrelaciones entre los términos utilizados para describirlos. Estas relaciones **son siempre recíprocas**.

Básicamente, existen tres tipos de relaciones: de equivalencia, jerárquicas y de asociación.

11.3.2.1. Relaciones de equivalencia.

La relación de equivalencia o "relación preferencial", se refiere a que los términos considerados **equivalentes** o sinónimos (de igual o muy similar significado), se agrupan en categorías de equivalencia, de manera que todos los términos equivalentes corresponden a un solo concepto unívoco. Esto significa que al momento de la búsqueda, se recuperan todos los documentos asociados a la categoría de equivalencia, aun cuando uno solo de los términos se utiliza como descriptor. Por ejemplo:



Aquí el término DIFICIL representa la categoría de equivalencia o **término preferencial**, al cual se asimilan los seis sinónimos señalados.

Es importante hacer una distinción entre **sinónimos** y **cuasi-sinónimos**. Los sinónimos son aquellos términos que tienen el mismo sentido en una disciplina específica. Los cuasi-sinónimos, en cambio, son términos cuyo significado puede ser diferente, pero

que **se hacen** equivalentes para las necesidades de un tesaurus en particular.

Esta relación de equivalencia o preferencial, que se señala por medio del símbolo USE, permite enviar desde un no descriptor (o término no usado) a uno o varios descriptores, con alguno de los siguientes propósitos:

- 1) para indicar un sinónimo preferido
ejemplo: Vejez USE: SENECTUD
- 2) para enviar desde un término específico a uno más general que se ha elegido para representarlo
ejemplo: Autobus USE: MOVILIZACION COLECTIVA
- 3) para desarrollar una abreviatura
ejemplo: KCI USE: CLORURO DE POTASIO
- 4) para definir una ortografía
ejemplo: Sicología USE: PSICOLOGIA
- 5) para recomendar el uso de uno o más descriptores para expresar un concepto
ejemplo: ADMINISTRACION DE BIBLIOTECAS PUBLICAS
USE: ADMINISTRACION
más BIBLIOTECA PUBLICA
- 6) para enviar de entradas indirectas a entradas directas, o viceversa, de acuerdo a las reglas de sintaxis del vocabulario
ejemplo: Funciones, tablas de
USE: TABLA DE FUNCIONES
- 7) para indicar terminología corriente
ejemplo: Pródromo USE: SINTOMA
- 8) para reemplazar términos corrientes por expresiones científicas más correctas
ejemplo: Dolor de cabeza USE: CEFALEA
- 9) para traducciones
ejemplo: Abstract USE: RESUMEN

INDIZACION COORDINADA

TG: INDIZACION

INDIZACION EN CADENA

TG: INDIZACION

11.3.2.3. Relaciones de asociación.

La relación de asociación (o de afinidad) se utiliza para indicar o representar aquellas relaciones entre los conceptos que, no siendo jerárquicas ni de equivalencia, parecen de interés o útiles de destacar.

Respecto a estas relaciones de asociación, y considerando que de hecho existe una gran cantidad de relaciones entre los conceptos, se recomienda ser conservador en este aspecto y establecer sólo aquéllas realmente útiles para los propósitos del análisis y la recuperación de información.

Para expresar esta relación asociativa se utiliza el símbolo TR (término relacionado).

Ejemplo: BIBLIOTECA

TR: CENTRO DE DOCUMENTACION

CENTRO DE DOCUMENTACION

TR: BIBLIOTECA

Las relaciones de asociación pueden usarse para expresar aspectos como los siguientes:

1) Antonimia.

DUREZA

TR: SUAVIDAD

2) Coordinación (términos derivados de un mismo concepto superior).

RELACION GENERICA

TR: RELACION PARTITIVA

3) Uso concurrente de dos términos.

EDUCACION

TR: ENSEÑANZA

4) Causa y efectos.

ENSEÑANZA

TR: APRENDIZAJE

5) Relación instrumental.

TRANSPORTE

TR: VEHICULO

6) Material.

LIBROS

TR: PAPEL

NEUMATICOS

TR: CAUCHO

11.3.3. NOTAS DE ALCANCE.

Una nota de alcance ("scope note") es una breve explicación del uso que se da a un descriptor, dentro del contexto de un tesoro en particular. No es una definición en el sentido que la entrega un diccionario. La Nota de Alcance, que se expresa con el símbolo NA, va inmediatamente después del descriptor.

Las notas de alcance suelen usarse para: 1) restringir el uso de un descriptor; 2) explicar abreviaturas y siglas; 3) excluir un posible significado.

Ejemplos:

ALEACIONES DE COBRE

NA: Usese para aleaciones en las cuales el cobre es el componente principal.

CENSO

NA: Enumeración oficial periódica de una población.

MAMIFEROS

NA: Sólo referido a las especies terrestres.

11.4. PRESENTACION Y ORDENACION DEL TESAURO.

11.4.1. LA INTRODUCCION.

Un tesoro debe incluir una parte introductoria lo más completa posible, donde se explique claramente su propósito, su estructura y el ámbito temático que cubre. Debe indicar, además, las reglas seguidas en su preparación, en qué obras se basó, metodología para la recopilación de los términos, etc. Un aspecto muy importante, también, es una explicación o guía de uso que incluya buenos ejemplos.

11.4.2. CUERPO PRINCIPAL.

Idealmente, un tesoro debe incluir una presentación sistemática y una alfabética de sus términos. La parte principal (o cuerpo principal) debe presentar la información completa sobre cada descriptor (en los que usan el sistema de términos preferenciales) o sobre cada concepto (en los que no usan términos preferenciales).

La parte principal puede estar ordenada en forma sistemática o alfabética, o una combinación de ambas. Cuando los descriptores y las permutaciones de los descriptores compuestos y los no descriptores se incluyen en una parte principal ordenada alfabéticamente, no se necesita un índice alfabético separado.

La información que suele incluirse en la parte principal en relación a cada descriptor es la siguiente:

- a) Representaciones del concepto.
 - número o notación del concepto
 - descriptor
 - no descriptor
 - referencia "usado por"

b) Información adicional.

- definiciones
- notas de alcance

c) Relaciones entre conceptos.

- conceptos dominantes (relaciones jerárquicas) TG
- conceptos genéricos (relaciones genéricas) TGG
- conceptos de entidad (relaciones partitivas) TGP
- conceptos subordinados (relaciones jerárquicas) TE
- conceptos específicos (relaciones genéricas) TEG
- conceptos parte (relaciones partitivas) TEP
- conceptos relacionados por asociación TR.

Ejemplos:

CIENCIAS DEL SUELO

- UP: Pedología
- NA: Se refiere a...
- TG: CIENCIAS DE LA TIERRA
- TE: FISICA DEL SUELO
QUIMICA DEL SUELO
- TR: ANALISIS DEL SUELO
CLASIFICACION DE SUELOS

14.03.01

ANTICONCEPCION/CONTRACEPTION

- NA: Acciones intentadas para impedir la concepción.
- TG: PLANIFICACION FAMILIAR
- TE: ANTICONCEPCION FEMENINA
ANTICONCEPCION MASCULINA
ANTICONCEPCION ORAL
ANTICONCEPCION REVERSIBLE
- TR: ABANDONO DE LA ANTICONCEPCION
ANTICONCEPTIVO
CAPACITACION EN ANTICONCEPCION
PREVALENCIA ANTICONCEPTIVA

11.4.3. PARTES AUXILIARES DEL TESAURO.

A fin de facilitar el acceso a la parte principal, los tesauros suelen contar con partes auxiliares, como índices alfabéticos, listados sistemáticos y representación gráfica.

Un **índice alfabético** es necesario cuando:

- la parte principal está ordenada sistemáticamente,
- la parte principal es una combinación de ordenación sistemática y alfabética,
- la parte principal sólo registra los descriptores.

Una **lista sistemática** es necesaria cuando:

- la parte principal está ordenada alfabéticamente
- la parte principal es una combinación de ordenación sistemática y alfabética.

Naturalmente, las partes auxiliares envían a la entrada apropiada en la parte principal, por lo que es útil asignar códigos o notaciones a los descriptores.

Los índices alfabéticos deben incluir descriptores y no descriptores. La permutación de los descriptores es muy útil para un mejor manejo de los descriptores compuestos. Esto se facilita ostensiblemente con un índice KWIC.

Ver ejemplos en la figura 44.

a) Ejemplo de un descriptor y sus relaciones.

08.09.03

COMERCIO EXTERIOR/ FOREIGN TRADE/ COMMERCE
EXTERIEUR

NA: Desde el punto de vista de cierto país o región.

TG: COMERCIO INTERNACIONAL 08.09.03

TE: EXPORTACION 08.09.03

TR: IMPORTACION 08.09.03

BALANZA COMERCIAL 08.09.03

SISTEMA DE CONTINGENTES 08.09.03

b) Ejemplo del índice alfabético de un tesoro en forma de índice KWIC.

EDUCACION SANITARIA	11.02.01
EDUCACION SEXUAL	10.06.00
EDUCACION	06.02.04
EDUCACION	09.05.02
EDUCACIONAL	09.05.01
EDUCACIONAL	06.02.02
EDUCADOR	06.02.05
EDUCATIVO	06.02.04
USE: MEDIOS DE ENSEÑANZA	
SISTEMA EDUCATIVO	06.02.02
USE: SISTEMA EDUCACIONAL	

c) Ejemplo de un índice jerárquico.

MIGRACION	15.01.01
· EXODO	15.04.01
· · EXODO INTELECTUAL	15.03.03
· · EXODO RURAL	15.02.03
· MIGRACION CIRCULAR	15.02.01
· MIGRACION COLECTIVA	15.03.02
· · MIGRACION FAMILIAR	15.03.02

Fuente: Tesoro Multilingüe sobre Población. París, CICRED.

Fig. 44: Partes de un tesoro.

Como sabemos, el papel esencial de cualquier servicio de información, es actuar como una interfaz entre una población determinada de usuarios y un universo de recursos de información ya sea impresos o en otras formas. Hasta aquí, hemos descrito diversas acciones que desarrollan estos servicios de información en la etapa que podríamos llamar de **entrada** o "input" al sistema, de acuerdo a lo que expresamos en el diagrama de la figura 1.

Siguiendo la idea del mismo diagrama, en la segunda parte de este libro nos referiremos a la **salida** o "output", que son las acciones orientadas a la recuperación de información. Al enfrentar esta segunda parte, es importante recordar una premisa que hemos planteado desde el inicio: el análisis y la recuperación de información son acciones totalmente relacionadas e interdependientes, y uno de los elementos que marcan esta interrelación es la utilización de un mismo vocabulario controlado en ambas etapas del trabajo documental.

SEGUNDA PARTE: RECUPERACION DE INFORMACION

"El objetivo de cualquier Sistema de Recuperación de Información es proporcionar a pedido, con máxima utilidad (recall) y precisión y con un mínimo de gasto, información pertinente, en respuesta a cualquier pregunta razonable, definiéndose como tal, cualquier pregunta seria de significación real o potencial, hecha por una persona que tiene una razón socialmente importante para desear la respuesta a su pregunta". (1)

(1) KENT, Allen. *Information Analysis and Retrieval*. New York, Becker and Hayes Inc., 1971 p, 147.