

## TEMA 1: LENGUAJES DOCUMENTALES

### *1.1: Concepto*

*Un lenguaje artificial que permite generar la representación formalizada de documentos y demandas relacionados con grupos de usuarios a fin de destacar los documentos que responden a las demandas.*

*Estas demandas son las que motivan la búsqueda de documentos en la descripción y la recuperación, que son los pasos básicos.*

*Estos pasos son de normalización ( se refiere a que está basado en normas) y normativo (que el mismo actúa como norma)*

### *1.2 : Funciones*

*La Cadena documental se compone de operaciones que básicamente son:*

- selección y adquisición
- análisis de los documentos: externo (catalogación)

*Interno (característica o indización)*

*descripción substancial o condensación*

- etapa de búsqueda o recuperación
- La difusión.

*El objeto básico de los lenguajes documentales es recuperar el documento en el menor tiempo con el menor esfuerzo.*

*Los lenguajes documentales cumplen dos funciones :*

- *Normalización: que se trata de la unificación de la terminología y de controlar los problemas de sinonimia y polisemia.*
- *Inducción: es la indicación del usuario de las posibles alternativas de recuperación de la información utilizando referencias cruzadas: ej; polución, vease contaminación)*

*Por tanto el lenguaje documental va a intervenir en la descripción del documento y en la recuperación de la búsqueda .*

- *descripción: una vez leido el documento se resume en elementos básicos que nos indica la materia de la obra, y después se traducirán esos conceptos bajo la normativa de la CDU.*
- *Recuperación: el usuario pide un libro y nosotros, traduciendo la demanda del usuario a los lenguajes documentales y por nuestra clasificación, recuperaremos el documento.*

*Por tanto los lenguajes documentales son los intermediarios entre los libros y los usuarios.*

- *Ordenación o archivo de documentos: por esquemas de clasificación (sea CDU por ejemplo) nos permitirá la ordenación y la recuperación del documento.*

### 1.3: Historia

Primero aparece la práctica de organización y mas tarde el concepto. En la edad media hay unos primeros intentos de la clasificación de materias y estos esquemas han ido avanzando progresivamente.

La aparición de la primera clasificación documental como tal se produce en el últimos cuarto del s XIX; antes de esto había un interés por la clasificación pero nadie se había parado a teorizar sobre el tema.

Dewey fue el primero que, siendo consciente y teorizando, escribió una obra.

De hecho, la CDU está basada en su obra. Pero el concepto de lenguaje documental no aparece hasta los años 50 del s. XX.

Así mismo, los Tesauros aparecen ligados a la aparición de este concepto y por tanto aparecen en la misma fecha.

Con la mayor automatización, la CDU sufren un retroceso utilizando más los tesauros ligados al ordenador. En 1959 aparece lo que se considera el 1º Tesauro, son los Tesauros vinculados a la de fensa de la EEUU.

## TEMA 2: TIPOLOGÍA DE LOS LENGUAJES DOCUMENTALES

### 2.1: Tipología de los lenguajes según su control terminológico.

Son los lenguajes – controlados

– libres

a) Los controlados son aquellos en los que se establece de antemano una lista cerrada y normativa de términos que pueden utilizarse de manera unívoca, es decir, se construyen a priori con un vocabulario predefinido, cuyos términos son los únicos aceptados. Son las clasificaciones, las listas de encabezamientos de materias, los tesauros, las listas de autoridades, etc...

Habrá que realizar una operación de catalogación y luego de traducción del lenguaje libre (el normal de andar por casa) a controlado. Los lenguajes controlados tienen unas ventajas: permiten una búsqueda rápida, más eficaz y más precisa, pero como desventajas está el hecho de que tienen una elaboración costosa, además necesitan una puesta el día constante. Otro problema será el de la ratificación internacional, para que una CDU salga se hará un estudio de las modernizaciones y luego que los comités de todos los países se pongan de acuerdo.

b) Los libres: no se define una lista de términos autorizados sino que se van extrayendo en lenguaje natural todos los conceptos tal y como aparece en los documentos, es un vocabulario no definido. Se van generando a medida que se produce el proceso de indexación. Las ventajas son la inversión de tiempo y de trabajo mínimos y el análisis es muy simple (no hay que traducir los términos) pero la búsqueda es menos eficaz porque no se tiene una lista controlada de términos.

Se van generando a medida que se produce el proceso de indexación.

Las ventajas : la inversión de tiempo y trabajo es muy poca y el análisis es simple (no hay que traducir los términos) pero la búsqueda es menos eficaz porque no tiene una lista controlada de términos.

## LISTA DE DESCRIPTORES LIBRES

## ***LIBRES***

### ***LISTA DE PALABRAS CLAVE***

*Esta tipología se establece mediante un criterio de coordinación de los lenguajes documentales .*

### ***PRECOORDINADOS***

#### ***DESCRIPTORES***

#### ***POSCOORDINADOS***

*A) precoordinados: son aquellos que coordinan los conceptos que constituyen una materia en el momento de la descripción del documento antes del almacenamiento de los documentos. Los términos están unidos por reglas sintácticas que establece el propio lenguaje . Son básicamente los sistemas de clasificación CDU y las listas de encabezamientos de matererias : Ej: Flora española en CDU es*

*581.9 es flora y 460 es España, por tanto los ° números se precoordinan para indicar que es flora de España 581.9 (460)*

*B) poscoordinados: yuxtaponen los conceptos en el momento del análisis y coordinarlos en el momento de la búsqueda y de la recuperación. No tienen sintaxis y normalmente su vocabulario está constituido por términos simples que se conectan en el momento de la recuperación por medio de una sintaxis booleana . Son las listas de palabras clave y los Tesauros (aunque estos últimos tienen cierto grado de coordinación) .*

*Ej: Flora, España (se yuxtaponen por medio de ; , : etc..)*

## ***2.2. ESTRUCTURA DEL LENGUAJE***

*Es la conexión y las relaciones que los términos del lenguaje tienen entre sí. Según esto los lenguajes documentales se suelen dividir en lenguajes según su estructura:*

### ***LENGUAJES DE ESTRUCTURA JERARQUICA***

### ***ESTRUCTURA LENGUAJES DE ESTRUCTURA ASOCIATIVA***

### ***LENGUAJES DE ESTRUCTURA SINTÁCTICA***

*a) Los lenguajes de estructura jerárquica o arbórea son los que organizan sistemáticamente el conocimiento, siguen un orden lógico agrupando conceptos más específicos dentro de los conceptos generales. En ellos los términos se relacionan por subordinación y las materias se agrupan en áreas categoriales, suelen ser los sistemas de clasificación CDU.*

*Ej: 3 – ciencias sociales*

*39 – folclore de las ciencias sociales*

*398 – folclore de las personas*

*398.3 – supersticiones*

- *Creencias y supersticiones relacionadas con las fiestas*

*Los tipos de lenguajes dentro de los lenguajes de estructura jerárquica o arborescente son:*

*a.1: clasificaciones enciclopédicas*

*a.2: clasificaciones facetadas*

*a.3: clasificaciones especializadas.*

*Otra forma de subdivisión mas correcta será:*

*a.1: lenguajes de estructura enumerativa*

*a.2: lenguajes de facetas*

*a.3: lenguajes mixtos o híbridos.*

*Estos lenguajes son precoordinados y sintéticos (que tratan de sintetizar la materia en un único concepto).los veremos después del punto c).*

*b) Asociativa o combinatoria: Son aquellos en que los términos que presentan los conceptos retenidos durante la indización se organizan alfabéticamente y se combinan entre sí sin quedar sujetos a una posición determinada del lenguaje, entre estos tenemos los encabezamientos de materias y los Tesauros. Presentan una lista facetada.*

*c)Lenguajes de estructura sintáctica: Son más elaborados que los que hemos visto y están constituidos por un conjunto de descriptores pero su gramática excluye cualquier ambigüedad posible a la hora de relacionar los términos.*

- *lenguajes de estructura jerárquica o arborescente*

*a.1 : Clasificaciones enumerativas:*

*Son el modelo puramente jerárquico. Son las más abundantes y corrientes por ser su construcción más sencilla.*

*Ejemplo: 398.332.32 (las de facetas : (44)19(038)*

*En las enumerativas solo ponemos eso, no especificamos mas.*

*a.2 : Clasificaciones facetadas*

*Todos los elementos se combinan. Cualquier documento puede clasificarse basándose en la combinación de facetas entre sí. Esto tiene su origen en dos figuras importantes:*

- *Ranganathan (el gran teórico)*
- *Bliss*

*La clasificación de Ranganathan es universal y por tanto más compleja ya que todos los términos son combinables entre sí. Se queda más en el campo teórico y no se trabaja con ella en el plano de la práctica. Es una clasificación facetada analítico – sintética.*

*a.3 : clasificaciones híbridas o mixtas*

Tienen su ejemplo más claro en la CDU. Tienen una parte jerárquica y otra facetada.

- lenguajes de estructura asociativa o combinatoria

b.1 : Lista de encabezamientos de materia

Son una correlación sucesiva de los encabezamientos que pueden expresar los temas de los cuales trata un documento. Es un lenguaje que requiere uniformidad y estricto control de terminología y creación de una lista previa. Establecen unas pocas relaciones entre términos. El lenguaje es precoordinado.

b.2 : Listas de descriptores

un descriptor es el conjunto de términos normalizados que expresan el contenido significativo de un documento. Estos pueden ser simples o compuestos:

- Simples: un único término
- Compuestos: varios términos o más de un sintagma.

Son elementos menos complejos y por ello se multiplican más a la hora de describir un mismo contenido. Son lenguajes poscoordinados.

b.3 : Los Tesauros

Comienza a utilizarse en los años 50. El primer tesauro como lo conocemos ahora surge en 1960 publicado por la Astia . Un tesauro es un vocabulario controlado y dinámico de términos que tienen entre sí relaciones semánticas, genéricas y que se aplica a un campo particular del conocimiento.

Los tesauros tienen que reflejar con precisión los términos y las referencias entre esos términos para poder realizar una buena indización. Va a tener habitualmente una lista jerárquica, alfabética..., una lista de grandes temas.

Tienen indices que nos permiten recuperar los descriptores, los más famosos indices son: Kwic, Kwoc.

El tesauro presenta relaciones semánticas de tres tipos:

- relaciones de equivalencia o preferenciales, nos remiten a otro USE UP
- relaciones jerárquicas (t.genérico y t. Específico) T.G Y T.E
- relaciones asociativas (términos relacionados) T.R