

TEMA 1.-

EL PAPEL DEL ANÁLISIS DOCUMENTAL DENTRO DE LAS CIENCIAS DE LA DOCUMENTACIÓN.

La documentación como tal surge con los teóricos. La documentación y sus teóricos surgen a finales del s. XIX y principios del s. XX, especialmente con L. Fontaine y Otlert. Serán los teóricos quienes organizan la documentación como disciplina y a través de unos instrumentos la convierten en ciencia, también serán los que la institucionalicen internacionalmente.

Son los creadores de la FID.

Son los teóricos también los que utilizan por primera vez las clasificaciones. Intentan crear un repertorio bibliográfico mundial pero sin mucho éxito, ya que era muy difícil.

Sin embargo, se dieron cuenta que la documentación estaba en la base de otras ciencias, es decir, que era una ciencia auxiliar de otras ciencias, dado que la documentación es la disciplina capaz de suministrar a otras ciencias la bibliografía necesaria para proporcionar el acceso a la información.

Es por lo tanto la documentación ciencia auxiliar de otras ciencias , que proporciona información a estas.

* En el desarrollo de la documentación hay tres etapas:

1ª Etapa: Se considera a la documentación como parte de la Biblioteconomía -> disciplina que clasifica las disciplinas por materias.

(Podemos apreciar que cada una de ellas tiene objetivos distintos, aunque parecidos. Por lo tanto se ve que la documentación es distinta en esta función a la Biblioteconomía).

2ª Etapa: La documentación se considera como la disciplina que recopila, interpreta y prepara los documentos.

(Aquí ya podemos observar como la documentación tiene un carácter instrumental).

3ª Etapa: Se le considera a la documentación como la ciencia que estudia y analiza la estructurar y las propiedades de la información, así, procesa, almacena y recupera la información. También disemina la información.

* Tres escuelas diferenciadas con respecto a la documentación:

1.- Escuela Anglosajona con Bradford y Taylor (a la cabeza).

Estas son las escuelas más avanzadas.

2.- La Escuela de la Europa Occidental

Es la escuela más nueva.

3.- La Escuela Soviética con Mijailov

En los años 60 surgen una serie de cambios:

1.- Surge la aparición de los ordenadores (en España es más tardío). Esto dota a la documentación de una transcendencia que no había tenido antes. Con ello surge también la creación de Bases de datos (aparición de Bases de Datos en CD Rom -> revolución en el almacenamiento y recuperación de la información).

El procesamiento de la información a través de los ordenadores supone un gran avance tanto en el almacenamiento, como en la recuperación de la información.

2.- Se empieza a concebir a la Documentación como información científica (es la que procesa, almacena, recupera, trata y difunde la información).

También en la década de los 60 surge una subdisciplina de la documentación que es la Bibliometría (sirve para cuantificar la información, quien, como, cuando, donde se produce esa información. Se utiliza en la planificación política de una país).

En los 70:

1.- En los 70 se introduce un concepto nuevo. Este concepto está muy vigente y se introduce en los EE.UU. como Information Science y se considera a la Documentación como ciencia que investiga las propiedades, comportamientos y medios de procesar la información, con el fin de que esta información sea lo más accesible posible. Comprende la producción, recogida, organización, elaboración y el uso de la información.

Este también es el concepto que tiene la FID hacia la documentación.

La documentación tiene como fin último SER ÚTIL, que se utilice. Por ello, el uso de la información es importante en esta disciplina ya que el fin último es que ésta información sea utilizada.

- Con la utilización se llega al concepto Anglosajón que lo puso de moda **Vickery** que es Información retrieval, es decir, recuperación de la información para que la utilice el usuario.

La información no tiene sentido sino es para recuperar la información, para que sea utilizada por el usuario.

* Papel que juega el Análisis Documental en Documentación

Courrier define al Análisis Documental como proceso de circulación de la información que tiende a establecer vínculos de comunicación entre el documento y el usuario. Es decir, el análisis documental son los medios técnicos y procesos que hacen posible la transferencia de información.

Hay autores como **López Yepes**, que dicen que el Análisis Documental es un auténtico proceso de información que tiene sus elementos propios (emisor, mensaje y receptor). Y su objetivo es al recuperación y la utilización de la información. Precisamente será la recuperación la que de al Análisis Documental un carácter distinto de cualquier otro elemento.

Para nosotros como estudiantes, el elemento principal es el mensaje de la información.

¿Qué debemos hacer para que el mensaje sea recibido y utilizado por el usuario?

* ¿Qué debemos hacer para que el mensaje sea recibido y utilizado por el usuario?

* Debemos representar el documento de una forma diferente a la de su estado natural para poderlo recuperar. Así, se procesa para su posterior recuperación y utilización.

Para representar el contenido de un documento se hace desde una actividad intelectual y analítica que permita

transformar ese documento. Es esto consiste el Análisis documental: analiza el documento y lo procesa para poderlo recuperar.

Ese proceso de transformación es el Análisis Documental. El proceso de transferencia de la información pasa también por una serie de etapas:

1.- Producción de la información: esto me origina los documentos primarios. (Este es el mensaje, es el documento con el que voy a actuar).

2.- La transformación o tratamiento de ese documento, que me da lugar al documento secundario.

3.- Almacenamiento de la información: por medio por ejemplo de una base de datos.

4.- Recuperación del documento. Que se puede hacer de dos formas:

– A petición del usuario

– Por DSI (Difusión Selectiva de la Información) –actualización de la documentación que se le va a suministrar a una persona cada x tiempo, se le suministra información dado que esta persona previamente la había solicitado–.

5.- Utilización de la información.

El Análisis Documental es fundamental en el tratamiento (en la transformación del documento) y en la recuperación de la información.

(Un documento bien indizado, es un documento bien recuperado y por el contrario, un documento mal indizado es un documento mal encontrado).

Por tanto:

El Análisis Documental es: Conjunto de operaciones necesarias para extraer la información o lo esencial de ella, contenida en los documentos primarios y prepararla para su posterior recuperación y utilización. Las herramientas necesarias para llevar estas tareas a cabo son los lenguajes documentales.

* Hay diferentes corrientes dentro de la concepción del Análisis Documental:

1.- Se considera el Análisis Documental sólo en la fase de recuperación. Representantes de esta corriente Coll-Vinent y Chaumier.

2.- Se considera el Análisis Documental en las dos fases: tratamiento y recuperación. Representante López Yepes.

* El Análisis Documental tiene varios niveles. Según la profundidad con que apliquemos ese análisis documental nos movemos en un nivel externo o en un nivel interno.

1.- Nivel externo:

Descripción física del documento. Es el análisis de los aspectos formales del documento.

La descripción física es unívoca, recoge los datos bibliográficos tal y como aparecen en el libro. Y estos datos bibliográficos sirven para identificar y analizar el documento. Es decir, es la catalogación bibliográfica, es el

asiento bibliográfico.

2.- Nivel interno:

Consiste en el análisis del contenido del documento. Este nivel es defendido por varios autores que dicen que en este análisis se debería incluir también el primero. Pero sobre todo este último nivel se ocupa del análisis del contenido, esto se hace a través de las clasificaciones o a través de las indizaciones. Este contenido se puede resumir mediante el resumen documental.

*Clasificación

*Indización nivel interno

* Resumen

Dentro de este nivel interno, hay diferentes corrientes:

2.1.- Corriente restrictiva: Esta corriente considera que solamente la indización y el resumen corresponden a este nivel interno.

Chaumier, Elisabeth Nett.

2.2.- Corriente integradora: Incluye a la clasificación dentro de este nivel.

PRACTICA 1.-

* Lista sistemática: parte del término cabecera y se van poniendo los del primer nivel y debajo de ellos sus específicos.

* Listas alfabéticas permutadas: aquella que alfabetiza las palabras no sólo las simples sino también las compuestas.

Ej: Atlas

Bibliografía

Catálogos bibliográficos

Catálogo – bibliográfico

* Clasificar jerárquicamente los siguientes términos:

Atlas

Catálogos bibliográficos

Obras de referencia

Enciclopedias

Bibliografías

Estadística

Diccionarios

Indicadores sociales

Directorios

Constituciones políticas

Indicadores económicos

Guías

Jurisprudencia

Legislación

Sentencias judiciales.

25 – Octubre – 1999

1.3.– NORMALIZACIÓN

Las primeras actividades relacionadas con la normalización comienzan en el siglo XVII vinculadas con intercambios comerciales y con la revolución industrial, por ejemplo, surgieron medidas estándares en los pesos, alturas, etc.

Las tareas de normalización han ido paralelas a los procesos de estandarización de la terminología.

En el siglo XIX con la revolución industrial se exige una normalización por el propio mercado.

En el siglo XX adquiere unos principios, que vamos a resumir en 5:

1.– La normalización es el acto que reduce la variedad en favor de la uniformidad.

2.– La normalización es una tarea de grupo, exige un consenso. La normalización no son leyes, pero se recomienda su uso.

3.– El valor de la normalización reside en su aplicación.

4.– La normalización debe tener una estabilidad en el tiempo. Nunca irá reñida con su revisabilidad, no deben ser eternas, hay que revisarlas según las exigencias, por ejemplo, homologación de los CD–Rom.

La normalización no son leyes, sino recomendaciones.

5.– Se debe tener en cuenta que la normalización es una actividad social que tiene unas repercusiones económicas, incidiendo pues en el mundo laboral

* En el campo del Análisis Documental, la normalización incide en tres niveles:

a) Selección

b) Tratamiento de la información, tanto en el nivel formal como en el nivel de contenido.

c) Difusión de la información (en soportes, etc...).

Definición de Normalización:

El concepto de Normalización ha sido definido por varias instituciones. Uno de los conceptos que mejor se ajusta al concepto de normalización en nuestro campo es:

Normalización: Forma de aplicar y establecer reglas con el fin de establecer un orden u homologación en cualquier campo de actividad con el consentimiento de todos los sectores implicados.

Definición según AFNOR con respecto a la documentación: La normalización es establecer reglas que aseguren la interconexión de sistemas y que faciliten el tratamiento y la transferencia de información.

Definición de Normalización Documental seguida por la UNESCO: Organización racional de los documentos y de sus soportes, que aseguren un intercambio eficaz de la información. Esta definición es la filosofía que ha seguido en Plan General de Información de la UNESCO.

* En el Análisis Documental la Normalización presenta dos actuaciones:

1.- La normalización de las operaciones propias del análisis

Indización bibliográfica,

Descripción

(Análisis) Resúmenes.

2.- La que se aplica a la recuperación de la información.

* Estrategia de búsqueda.

* La de los propios lenguajes de recuperación de las Bases de Datos.

* El grado de incidencia de la normalización es distinto según el nivel en el que se aplique.

Respecto al análisis formal:

– Si lo que tengo que normalizar es la descripción física de un documento, es más fácil, los organismos oficiales han llegado a un consenso respecto a las reglas a seguir que se han fijado.

– Hay más problemas en el análisis de contenido (análisis interno). Es una tarea más complicada, no hay tantas reglas. No hay consenso.

En definitiva: Es más difícil la normalización en el nivel de contenido de los documentos, que en el nivel externo.

! Respecto al nivel externo

– Ventajas: Una gran ventaja es el intercambio de información.

Ejemplo: el OPAC. En las referencias que tiene el OPAC hay normalización.

– Normas: ISO 690/1987

AACR (Anglo American Cataloguing Rules)

(Reglas de catalogación Angloamericanas)

Reglas o normas de catalogación españolas

ISBD (International standard bibliographic description)

ISO 2709 / 1981 (desarrolla el formato MARC)

ISBN

NIPO (Código para identificar documentos).

! Normalización respecto al nivel interno.

Como en el nivel interno están implicados conceptos intelectuales, es más difícil de normalizar estos procesos, por lo que siempre será subjetivo y esto está reñido con la normalización.

En el análisis de contenido existen dos procesos:

a) sintético: clasificaciones que datan desde la antigüedad. Clasificaciones que existen desde hace mucho tiempo. Siempre ha habido un afán de clasificar desde siempre el saber humano.

b) analítico: indización y resumen.

Las clasificaciones están muy normalizadas (ejemplo, con una serie de dígitos) y existen muchas que se utilizan en Bases de datos.

Hay dificultades para normalizar el análisis de contenido:

Dificultades:

* Una de las dificultades es de **tipo lingüístico**, la solución o la forma idónea es establecer lenguajes documentales, pero estos topan siempre con la dificultad de los conceptos. No es fácil encontrar términos que definan exactamente algunos términos. Es decir, el concepto de droga no es lo mismo para un médico, que para mí.

El Análisis Documental ha desarrollado algunas herramientas para subsanar el problema lingüístico, a través de por ejemplo los TESAUROS (con ellos consigo ver que término corresponde a un concepto concreto. Además evito la polisemia y la sinonimia, aunque siempre existen las falsas combinaciones de términos, por ejemplo: documentación de la historia frente a historia de la documentación. Es decir, vemos que siempre en un lenguaje de recuperación la combinación de términos, es decir, juntar dos palabras puede dar un resultado que no queríamos.

* Existe sin embargo otro **problema** en la actividad analítica, esta vez **generado por el propio sistema documental** tanto de indización como de recuperación, que implica restricciones a la hora de la recuperación por el ruido y el silencio debido a las dificultades de los lenguajes documentales.

Aquí también entra el problema de las propias limitaciones del documentalista. El juicio del documentalista es subjetivo. El tiene su propia cultura, lengua, etc.. incluso la propia consistencia de la documentación.

La normalización en el nivel interno o en el nivel de contenido tiene que ir orientada en dos sentidos para evitar estos problemas:

Es decir

- los problemas de tipo Lingüístico
- y los problemas del propio análisis.

1.– Redactar o realizar unas recomendaciones de las metodología a seguir en las tareas tanto de indización como de resumen. (No se puede indizar ni resumir si previamente los documentalistas que van a realizar esa tarea no se reúnen y dictan unas normas). Aunque existen unas normas, estas no son de obligado cumplimiento, por lo que algunas se pueden modificar, por ello se debe establecer la metodología a seguir.

Las reuniones han de ser periódicas para establecer una metodología flexible, modificaciones, etc.... Hay que ver también en esas reuniones si se llega a algún resultado o no. En todo ello debe haber un consenso.

2.– Elaborar normas a seguir en los lenguajes documentales (sobre todo en los tesauros).

Sutter dice que la normalización del análisis de contenido ha de seguir de cerca las propias técnicas documentales (La normalización ha de tener en cuenta los cambios de estas técnicas documentales).

La estabilidad de la norma tiene que contemplar los cambios y la evolución de la sociedad.

* En cuento a la normalización en la indización.

– Normas de Indización:

ISO 5963 / 1985 (establece los métodos para el examen de documentos. Se explica como determinar el contenido y los términos de indización).

ANSI (American National Standar Institu)

UNESCO (Principios de indización).

AFNOR (Principios de indización)

Harold BORKO (Autor que estudia el control de la normalización en los lenguajes de indización).

ERIC (Instituto General de Educación. Este también ha echo unas normas para normalizar los lenguajes documentales).

Normalización en el resumen

ISO 214 / 1976

ANSI

CHEMICAL ABSTRACT

UNESCO

EVINITI

Harolg BRONKO

Normativa en cuanto a los TESAUROS

ANSI (Edición de guía para el mantenimiento y creación de tesauros).

FID

UNESCO

ISO 5964 / 1985

AFNOR

La forma de producir documentos hasta hace poco no estaba normalizada. Por lo que las normas que eran imprescindibles para crear una Base de datos han ido retrasadas.

El que se normalice el nombre del autor es muy importante, dado que el autor firma como le da la gana (él no está obligado a firmar de una manera determinada: nombre, nombre y apellidos, etc...). El nombre del autor no se puede homologar, hay que ponerlo siempre como aparece. La homologación del propio documento tiene tanta importancia como la del contenido. El cómo publicar está homologado.

Ventajas de la normalización en el Análisis Documental.

- 1.- El documentalista se encuentra una estructura ordenada presentándola por ello con una mayor eficacia, y facilitando la labor.
- 2.- Se facilita el almacenamiento de la información.
- 3.- Los productos finales de la indización del análisis del contenido son más precisos. Estos productos finales son los Índices. (Qwick)
- 4.- Ahorra tiempo y rapidez en cuanto a la elaboración de documentos secundarios. (Qwock)
- 5.- Se facilitan las tareas al usuario.
- 6.- Difusión internacional de la información.

* Es importante también normalizar los propios canales de difusión de la información. La standarización. Es decir, homologar boletines, índices, normalizar documentos secundarios, los soportes (en los que está almacenada la información), la difusión selectiva de la información DSI, las propias difusiones, etc...

Proceso que se sigue para hacer una norma. Organismos de la normalización.

Los organismos de normalización pertenecen a dos ámbitos geográficos:

- Nacional: Oficinas y Servicios de información AENOR .

– Internacional: La incidencia mayor la tienen los organismos internacionales. Norma ISO (reconocida en el mundo entero).

* Luego las disciplinas de carácter sectorial:

IFLA ----- Bibliotecas CIA ----- en archivística.

FID ----- en documentación

2.- PRÁCTICA

Ordenar de forma jerárquica, señalando los términos genéricos, términos específicos y términos relacionados.

1º Transporte

transporte transporte de cercanías

modos de transporte transporte aéreo

aviones transporte motorizado

carros puertos

barcos metro

aeropuertos estación de mercancías

parada de autobús cobertura del transporte

transporte internacional transporte nacional

helicópteros vías navegables

estaciones de trenes helipuertos

microbuses transporte interurbano

líneas aéreas canales

trenes aeródromo

transporte urbano taxis

submarinos transporte fluvial

tranvías estaciones de viajeros

automóviles transporte marítimo

terminales de transporte

8 – Noviembre – 1999

Organismos que se encargan de la normalización:

Hay dos niveles de organismos:

! Organismos nacionales:

- Oficinas y servicios de normalización. Están especializados según el área a la que destinan sus normas.
- AFNOR, y AENOR ; coordinan y difunden, se encargan de centralizar todas las normas que emite la Nación.
- Autoridad (Estatual o no), que se encarga de supervisar la aplicación y puesta en práctica de todas las normas.

! Organismos en el ámbito Internacional:

- IFLA (Bibliotecas)
- CID (Documentación)
- ISO ; tiene dos ámbitos de aplicación.

*Uno el geográfico; en cuanto a su extensión geográfica.

*Y otro en cuanto a las materias, ya que normaliza todos los campos.

El objetivo principal: favorecer la normalización, esto lo va a llevar a cabo a través de los Comités técnicos (TC), que están compuestos por varios comités, y a su vez por varios subcomités, etc..., y también por una serie de observadores.

La misión de los Comités Técnicos (TC) , son varias:

- Elaborar normas nuevas, recomendaciones y revisiones de normas antiguas.
- Intercambio de información entre los organismos miembros.
- Difusión de documento, es decir, que estos documentos sean emitidos.
- Una de las misiones y normas que más nos atañe a nosotros es el ISO / 46, que a su vez se compone de 25 organismos participantes y aproximadamente una veintena de organismos observadores y varios subcomités.

La ISO además de estos comités técnicos tiene un centro de documentación en Ginebra y un comité permanente encargado de estudiar la información científica técnica.

La Elaboración de las normas pasa por una serie de pasos:

- 1.– Inclusión del tema a tratar, en cada comité técnico, en el programa.
- 2.– Estudio del anteproyecto
- 3.– Preparación del proyecto como norma y la votación por los comités miembros.

- 4.- Pasa obligatoriamente por la aprobación de al menos el 75% de los miembros del comité.
- 5.- Examen definitivo del proyecto.
- 6.- Aprobación definitiva del proyecto (que ya es norma), y su publicación como tal norma internacional.

Estas seis etapas corresponden a tres fases distintas:

a) El desarrollo de los expertos:

Esta primera fase corresponde con las etapas 1 y 2. La primera de estas etapas viene a tardar un periodo de 12 a 24 meses, y en la segunda se consideran aproximadamente unos 20 meses.

b) La segunda fase: Procedimiento de aceptación:

Corresponde a las etapas 3,4, y 5, aquí se hace la puesta a punto, que tarda aproximadamente unos 20 meses.

c) La tercera fase: Definitiva impresión y publicación de la norma:

Corresponde con la 6 etapa. La impresión y publicación de la norma se suele hacer al mes siguiente de haber superado las otras.

Se exige una perenne vigilancia en el tiempo, pues se debe revisar continuamente la norma por la tecnología nueva que se va implantando. Se debe también tener en cuenta la flexibilidad de la norma. La normalización de ISO choca con características constitucionarias (de usos y costumbres del país). Los comités de ISO deben saber plasmar que esa norma que están emitiendo debe poderse adaptar a las características propias de ese país donde se va a adoptar.

Se recomienda que la norma sea revisada cada 5 años.

Existe un catálogo de normas que cada comité ha dictado con anterioridad, para esto cuenta con un centro de documentación.

La utilización de las normas no es obligatoria (no es de obligado cumplimiento en el sentido de que no esté legalizado el no cumplimiento). Pueden ser obligatorias en ciertos casos excepcionales.

Se recomienda que sean utilizadas por el mayor número posible. No obstante, existe sensibilidad, hay un consenso en cuanto a las normas.

CRITERIOS DE EFICACIA DE UNA NORMA

Los criterios de eficacia de una norma se pueden mediar por las siguiente cualidades de esa norma:

- 1.- Por la capacidad respecto a la necesidad que ha motivado la emisión de esa norma.
- 2.- Por la facilidad del empleo.
- 3.- Que contenga unas instrucciones claras y precisas.
- 4.- Que sea fácilmente utilizada por el usuario.
- 5.- Es necesario que la norma que se emita tenga un campo de aplicación que esté aceptado por las distintas

lenguas, razas, etc, es decir, por otras culturas.

15 /11 / 99

TEMA 2.- LENGUAJES DOCUMENTALES

2.1.- Lenguajes Documentales:

El tratamiento documental engloba dos aspectos:

- a.- Hacer el análisis documental
- b.- Recuperación de la información

Estos dos aspectos están ligados, para que exista el nexo de unión entre ellos necesitamos los lenguajes documentales. Para que los lenguajes documentales sean efectivos se tiene que utilizar el lenguaje natural, pero tendrá que ser normalizado.

Los lenguajes documentales se utilizarán indistintamente en el análisis documental y en la recuperación de la información.

Es imprescindible en el tratamiento documental tener un lenguaje de documentación preestablecido.

Los lenguajes documentales surgen en el siglo XIX al tiempo que nacen las primeras clasificaciones documentales. En el siglo XX llegan a su máxima exposición, comenzando con Cutter.

Las nuevas tendencias en los lenguajes documentales van encaminadas al tratamiento documental a texto completo.

2.2.- Definición de LENGUAJES DOCUMENTALES

a.- Según VAN SLYPE, Lenguajes Documentales es todo sistema de signos que permiten representar el contenido de los documentos con el fin de recuperar a los que sean pertinentes, no se refiere a los criterios utilizados en la búsqueda documental tales como autor, lengua, etc...

Hay que resaltar el concepto de PERTINENCIA, que se conseguirá a través de un buen control del vocabulario.

b.- Según Banca Gil Urdiciain los Lenguajes documentales son sistemas artificiales de signos normalizados que facilitan la representación formalizada del contenido de los documentos para permitir la recuperación, ya sea esta manual o automatizada, de la información solicitada por los usuarios.

Hay que destacar sistema artificial de signos. Artificial porque es opuesto al natural, ya que los lenguajes documentales utilizan la palabra normalizada, dotan a la palabra normalizada de un único significado. El lenguaje natural se utiliza de forma mediata. El lenguaje artificial es estereotipado y unívoco, no es arbitrario, pasa por controles de normalización, control léxico, etc...

Características de los LENGUAJES DOCUMENTALES:

a.- Relación con otras disciplinas: lingüística documental, la lógica, semántica, inteligencia artificial, matemáticas, lingüística textual (aparece en los años 60, estudia el contenido, va más allá y se preocupa por el contenido de los documentos).

b.- Lenguaje no natural sino artificial, aunque utiliza los signos/símbolos del natural, pero estos términos del lenguaje natural adquieren un valor semántico (UNÍVOCO) por medio de reglas morfosintácticas y la normalización. El signo son las palabras.

Principios de los LENGUAJES DOCUMENTALES.

1.- Es un lenguajes que no puede ser REDUNDANTE, cuanto más se elimine la redundancia más aumenta la eficacia.

2.- Se debe tender al principio de ENTROPIA, el ahorro de técnicas. Debemos expresar el mayor número de conceptos posibles con el menor número de términos posibles, sin perder información. Esto es una distinción con el lenguaje natural, volviendo al concepto de univocidad. A un mismo concepto le corresponde un término y un mismo término le corresponde un mismo concepto.

3.- Eliminación de la AMBIGÜEDAD, un término con diferentes significados crea confusión.

2.3.- Objetivos y funciones de los LENGUAJES DOCUMENTALES

Los Lenguajes Documentales tienen dos objetivos principales:

a.- Hacer el análisis documental:

Este objetivo persigue la normalización, y quiere decir la capacidad de representar el contenido de un documento determinado por medio de esa normalización.

Este primer objetivo lo que hace es que si yo normalizo el vocabulario lo que hago es reducir considerablemente el volumen de términos a manejar en ese vocabulario controlado.

Una de las causas por las que reduzco el vocabulario es porque en el lenguaje documental:

- se queda en los sintagmas nominales o sustantivos y elimina los verbos.
- Ante dos palabras similares se debe elegir una de ellas, para evitar la ambigüedad.

El tamaño idóneo. ¿A qué volumen se debe tender?. No hay una respuesta específica. Va a depender de la temática de la que trate. Por ejemplo el vocabulario es más controlado en ciencias experimentales y son más voluminosos que en las áreas sociales y humanidades. Por eso hay más tesauros en estas áreas que en ciencias experimentales.

b.- Con respecto a la recuperación

Va dirigido a la inducción, es decir, que los Lenguajes Documentales son capaces de suministrar el vocabulario exacto de los conceptos que no ha exigido la demanda de información. Y además, es capaz de encontrarme esos términos a nivel de especificidad y generalidad que el usuario nos demanda.

En un Lenguaje documental se reduce y prácticamente se elimina el problema del multilingüismos. Ya la mayoría de los tesauros son multilingües, este multilingüismo también se reduce en las clasificaciones

PRÁCTICA 3

TRANSPORTE

TG. Modos de transporte (usado por medio de transporte)

- Transporte aéreo
- Transporte no motorizado
- Transporte fluvial
- Transporte marítimo

* Puerto

TG. Cobertura del transporte

- Transporte internacional
- Transporte nacional
- Transporte interurbano
- Transporte urbano
- Transporte de cercanías.

TG. Terminales de transporte

- Aeropuertos
- * Aeródromo (como sinónimo)– Base aérea
- Estaciones
- * Estación de mercancías
- * Estación de viajeros
- Parada de autobús – estación de autobús USE
- Estación de tren
- * Helipuerto
- * Puerto

Elementos que constituyen los lenguajes documentales:

* **Descriptores:**

Palabras que sirven para describir la información, por medio de estas constituimos los descriptores . Las palabras se toman del Lenguaje Natural pero para formar los descriptores se reduce a una forma gramatical uniforme e invariable -> generalmente suelen ser en sustantivo en singular, pueden ser simples o compuestos.

***No Descriptores:**

Las palabras del lenguaje natural que tienen relación con los descriptores pero, se inscriben en unas listas anejas que nos remiten continuamente a los descriptores las denominamos no descriptores, forman un listado que se puede recuperar y nos envían a los descriptores.

Son palabras controladas por el lenguaje documental, no se pueden utilizar para describir la información.

*** Relaciones entre descriptores:**

Existen relaciones jerárquicas, de equivalencia y de asociación, esto nos permite agrupar los distintos conceptos dentro de un mismo término, nos sirve para precisar la búsqueda o extenderla. Son objeto de normalización y se indican con un código.

*** Notaciones numéricas, notaciones alfabéticas o notaciones alfanuméricas:**

Se trata de un código de números o letras o combinación de ambas que sirve para identificar los descriptores y hacerlos figurar de esta forma abreviada tanto en las reseñas bibliográficas como en los ficheros.

*** Notaciones generales:**

De aclaración de aplicación, etc... Precisan el sentido en que debe emplearse el elemento. Pueden considerarse como una especie de definición.

***Elementos sintácticos que utilizamos con los descriptores.**

Permiten relacionar los descriptores entre sí o indicar su función.

2.3.- Tipología de los LENGUAJES DOCUMENTALES

Los Tipos de lenguajes documentales van a depender de los distintos autores y de las distintas escuelas.

! Para clasificar los Lenguajes documentales existen diferentes criterios, nosotros vamos a seguir los siguientes:

- Los criterios más generales son el de coordinación de los términos
- El control de los términos
- La estructura de los términos
- El nivel de análisis de los términos.

A.- Coordinación de los términos

Teniendo en cuenta el criterio de coordinación de los términos tenemos:

Lenguajes precoordinados (dentro de estos tenemos:

- clasificaciones
- encabezamientos de materia)

Lenguajes postcoordinados (donde tenemos:

- listas de descriptores libres
- listas de palabras clave
- tesauros)

Con respecto a esta tipología, el criterio de coordinación es aquel que tiene en cuenta la combinación de los términos, si esta se realiza a posteriori es decir, antes de la indización o antes de la recuperación estamos hablando de un lenguaje postcoordinado. Si es después, hablamos de un lenguaje precoordinado.

* Combinación: cuando yo para recuperar información tengo que combinar los términos, esta forma es la que me dice si es un lenguaje precoordinado o un lenguaje postcoordinado.

Ejemplo:

Ponemos por un lado **cáncer**, y por otro lado tenemos **prevención**

(para recuperar la información tengo que combinar los dos términos. Estamos hablando de un lenguaje postcoordinado)

Si lo que tengo es **prevención del cáncer**, estamos hablando de un lenguaje precoordinado ya que se realiza a priori esta combinación de los términos.

Otro ejemplo:

En las clasificaciones tengo ya una lista establecida, entonces hablamos de precoordinado.

Qué ocurre con los lenguajes precoordinados, estos son los primeros que surgen en los índices impresos.

Los lenguajes de recuperación eran precoordinados.

En las búsquedas manuales son más efectivos los lenguajes precoordinados, aquí yo leo una cadena de caracteres y yo identifico a través de esa cadena de caracteres.

Más tarde, los índices impresos se informatizan, por lo que puedo recuperar la información con operadores booleanos, con lo cual surgen los lenguajes postcoordinados.

En las nuevas bases de datos conviven conjuntamente lenguajes precoordinados con lenguajes postcoordinados.

* Ventajas de los Lenguajes Precoordinados

- recuperan con mayor especificidad.

* Ventajas de los Lenguajes Postcoordinados

- al hacer combinaciones recupera con mayor exhaustividad, el peligro más grave de los postcoordinados es la falsa combinación.
- También pueden aparecer ruidos y silencios.

Van Stype: dice para resumir la precoordinación y la postcoordinación:

Los Lenguajes precoordinados son los lenguajes de clasificatorios o sintéticos.

Y los Lenguajes Postcoordinados son los lenguajes de indización o los analíticos.

En los sistemas precoordinados la coordinación de términos sigue unas reglas sintácticas– tienen unas relaciones gramaticales.

Los lenguajes precoordinados son los que habitualmente se han utilizado en bibliotecas.

En cambio, los lenguajes postcoordinados – la relación entre los términos no es gramatical sino lógica, es decir, carecen de sintaxis. Se pueden recupera por medio de operadores booleanos.

Ejemplo:

Un documento que trata sobre el impuesto sobre la renta de las personas en el País Vasco.

Indizaría con descriptores de este tipo: Política, económica, impuesto, País Vaco.

B.–Con respecto al control de términos:

CONTROL	Lenguajes libres	– Palabras clave – Descriptores libres
	Lenguajes controlados	– Clasificaciones – Listas de encabezamientos de materia. – Tesauros

Dependiendo del control que ejerzamos sobre el vocabulario a usar, el **lenguaje libre** utiliza el lenguaje natural, pero que presenta dificultades a la hora del tratamiento de la información. Se componen de un vocabulario de palabras naturales.

El propio ordenador extrae palabras significativas y con ellas elabora ese lenguaje libre.

Algo que caracteriza los lenguajes libres es su gran ambigüedad semántica, no elimina sinónimos, polisemia, etc...

Utiliza la información que suministra el propio autor del documento. Da mucho mejor resultado en Ciencias Experimentales que en Ciencias Sociales, es necesario que los usuarios conozcan el lenguaje porque sino es más difícil recuperar, necesitan pues conocer las expresiones utilizadas con anterioridad para su recuperación.

– Es difícil la recuperación del lenguaje libre, sin embargo es más fácil la indización.

Si se quiere recuperar con mucha precisión se debe recurrir al Lenguaje controlado. Y si deseo exhaustividad se debe recurrir al L. Libre.

Las entradas de los lenguajes libres y controlados forman una lista de palabras que yo puede utilizar para recuperar la información.

Las entradas de los lenguajes libres son una lista de palabras ordenadas alfabéticamente.

Con respecto a los **Lenguajes controlados**, estos me definen todos los términos que yo debo utilizar para representar un concepto. De estos términos se utiliza uno sólo, pero sí existen reenvíos (con el lenguaje libre tendríamos los dos términos del normalizado y el reenvío).

– Así como los lenguajes libres son muy extenso, los Lenguajes Controlados son más reducidos, es más simplificado (dado que yo para decir algo utilizo en vez de tres palabras, una). A penas utiliza enlaces sintácticos, gramaticales, y además son más manejables.

El lenguaje controlado suple deficiencias del L. Libre y el Libre suple deficiencias del controlado.

C.– **En cuanto a la estructura:**

	– JERÁRQUICA	–clasificaciones – léxicos (palabras clave y descriptores libres)
ESTRUCTURA	– COMBINATORIA	– Tesauros
	– SINTÁCTICA	– Gramática simple (encabezamientos) – Gramática compleja (Resumen documental)

C.1.– **Los lenguajes en cuanto a la estructura jerárquica:**

Cuyo ejemplo son las clasificaciones.

Las clasificaciones organizan el conocimiento sistemáticamente. Agrupan las materias en áreas por categorías.

La representación de estos conceptos en las clasificaciones pueden ser alfabéticas, numéricas o alfanuméricas.

Tienen una estructura arborescente, parten de su genérico a su específico. Son categorías inamovibles y sirven más que para indizar un documento para localizar el documento. Uno de sus mayores inconvenientes es la rigidez. Son inamovibles. Son muy rígidos. La rigidez de las clasificaciones impide una puesta al día. No se utilizan para indizar. Las clasificaciones se usa mucho en bibliotecas.

Existen clasificaciones especializadas sobre un campo determinado del saber, se ajustan mejor al documento.

Las clasificaciones a facetas (Colón) dividen la información en categorías principales que a su vez se divide en subcategorías que se combinan entre sí.

C.2.– **Según se estructura combinatoria**

Los lenguajes de estructura combinatoria permiten la libre coordinación de los términos según las necesidades de la propia indización.

El tesauro además de combinatorio es jerárquico.

C.3.– **Según su estructura sintáctica**

Los lenguajes que pertenecen a la estructura sintáctica recurren a la sintaxis para poner en relación o combinar los distintos elementos, es decir, son unas veces como frases enteras.

D.– Según el nivel de análisis:

Superficial	Clasificaciones
Profundo	– resúmenes documentales – tesauros

Este es el menos utilizado, de hecho hay muchos autores que no la consideran.

Según gran parte de ellos los lenguajes documentales se pueden resumir en dos:

- Lenguajes de clasificación
- Lenguajes de indización.

Los lenguajes de clasificación no son motivo de estudio en esta asignatura, pero sí la indización.

Los lenguajes documentales más usados en España son:

- Listas de encabezamientos de materia.
- Tesauros
- CDU

¿ De todos los diferentes tipos de lenguajes, se describe brevemente cada uno de ellos:

! Listas de encabezamientos de materia:

Tratan de representar uno o todos los temas de un documento y se sirven de la ayuda de una serie de términos expresados en el lenguaje natural y entre ellos hay una relación. Son por tanto precoordinados. Se componen de encabezamientos y subencabezamientos los más utilizados en España son los de la Library of congress – Bilindes.

! Tesauros:

Lenguaje postcoordinado de estructura combinatoria, compuesto por una lista estructurada de conceptos que posibilitan la descripción analítica del contenido de los documentos. Los tesauros implican un altísimo control de los términos. Si los términos se combinan la precisión es mayor. Son la herramienta que mejor soluciona los problemas del lenguaje natura. Se evitan los problemas de sinonimia, polisemia, suelen estar formados por unitérminos, esto no evita que haya términos compuestos.

! Clasificaciones

Es un lenguaje precoordinado, controlado y de estructura jerárquica. Se usa mucho en bibliotecas.

! Listas de palabras clave

Son listas no estructuradas (se ordenan alfabéticamente) de palabras significativas o no vacías, seleccionadas por medio del ordenador a partir del título, resumen o texto completo, para describir el contenido de los documentos. Es única para cada documentos. Es un lenguaje postcoordinado y de estructura libre.

! Lista de descriptores libre

Son listas no estructuradas (se ordenan sólo alfabéticamente) de conceptos destacados a partir de los documentos que se analizan. Se expresan o bien por palabras o bien son expresiones que se extraen por medio del documento o a propuesta del propio indizador.

La existencia de uno u otro descriptor no se verifica en una primera lista, se puede ir metiendo poco a poco.

Los descriptores libres surgen en contraposición a los encabezamientos de materia.

TAUBE sustituye los encabezamientos de materia por los descriptores libres.

Pertenecen a la tipología de postcoordinados.

(Yo puedo elegir ante una demanda concreta el descriptor que quiero).

! Lenguajes de estructura sintáctica:

Que se forman a partir de términos que son susceptibles de ser combinados. La combinación se hace a partir de una sintaxis más compleja, se fundamentan en reglas gramaticales que pueden ir desde un simple indicador de Rol, hasta el establecimiento de nexos de unión que asocian los diferentes conceptos de la frase.

Como característica cabría destacar la rigidez de estos lenguajes documentales. Son lenguajes biunívocos, sólo hay un término para determina un solo objeto. Cada término no designa más que un sólo objeto.

22 / 11 / 99

TEMA 3.-

3.1.- TERMINOLOGÍA Y DOCUMETACIÓN

La terminología es una disciplina con unos siglos importantes en el análisis. Es una disciplina muy joven, empieza su andadura por los años 30, es multidisciplinar, tiene conexiones con lógica, semántica, etc...

Definición breve de terminología:

Conjunto de vocablos propios de determinada profesión, ciencia o materia.

También se la define como arte de coleccionar todas las palabras de un idioma.

El resultado de la terminología son unas herramientas imprescindibles, ya que son diccionarios, tesauros, ... (es imprescindible para ordenar conceptos profesionales y científicos y transmitir el conocimiento de una lengua a otra a través de mediadores lingüísticos). También almacena y recupera información.

Por lo tanto, el documentalista debe tener nociones de terminología.

A través de la aplicación de la terminología surgen lo que denominamos **Lenguajes profesionales**, esta es la jerga de una profesión, se aprende practicando constantemente. En los Lenguajes Documentales se exige una

normalización, no sólo de concepto sino también del término, con los lenguajes profesionales ocurre lo mismo que con el lenguaje natural exige una cierta normalización, pero también una rigidez relativa.

La evolución debe ser controlada por la terminología, por lo que debe ser flexible.

También ocurre con los Lenguaje Profesionales que hay términos que se acuñan de lenguas extranjeras que no existe el término en Español, por lo tanto, se adopta el término en esta lengua extranjera (casi siempre término anglosajón) y de alguna manera, se adapta a nuestro idioma, bien adaptándolo o bien transcribiéndolo a nuestra lengua. Ej. Fútbol.

De todo esto, se ocupa la terminología.

La terminología debe planificar la comunicación profesional en su lengua materna, es decir, decir las cosas de forma clara.

El término en terminología es algo más que una palabra, puesto que el término en terminología debe expresar un concepto completo.

• Cómo se forman los términos.

Se forman de la misma manera que en lenguaje natural, pero en terminología surgen de forma planificada para el lenguaje profesional.

• Los elementos de un término son los siguientes:

- Terminologización :

La palabra adopta en su valor semántico previo un significado específico, y este significado debe corresponder a un determinado concepto de una determinada rama del saber.

Ej: Restaurar: todos sabemos lo que es, pero en historia tiene una valor especial, nos referimos a Hª de la restauración.

- Composición

Consiste en la agrupación de diversos elementos. Ej: Maquina de escribir.

- Derivación

Consiste en que los elementos del término varían según la lengua. Algunos son internacionales como la palabra Pseudo.

4. Conversión

Son nombres que no cambian de clase. Estamos en la rama del saber que estemos siempre significan lo mismo. Ej. Hercio, Diesel,.....

5.- Préstamo

Adaptación de un término en el idioma concreto del país donde se esta hablando al nuestro.

Ej: fútbol

Hay términos que se adoptan con una escritura adaptada a nuestro lenguaje, el ejemplo de fútbol es también representativo.

6.- Abreviatura o siglas

Que se convierten en el lenguaje profesional aceptado. Ejemplos: OTAN, ONU, ...

Los sujetos que llevan a cabo la terminología son expertos en las diferentes ramas del saber. Es fácil deducir quién es el sujeto que realiza la terminología -> los especialistas.

Pero también es necesario saber quién son los usuarios de la terminología:

- Traductores
- Cuerpo docente y discente (profesores y alumnos)
- Centros de información y documentación

En este caso, ¿para qué se usa? Para varias de sus actividades tales como:

- 1.- Elaboración de tesauros
- 2.- Elaboración de repertorios y bases de datos
- 3.- Elaboración de léxicos (tanto monolingües como multilingües). Un diccionario multilingüe nos ayuda para saber cual de las palabras sale de nuestro propio país, etc...
- 4.- Asignación de términos: para saber con qué términos hemos de asignar conceptos nuevos que aparezcan, es decir, para indizar.
- 5.- Recuperación de información: A la hora de expresar una búsqueda concreta y a la hora de interpretar lo que un usuario nos demanda.

• Relación concreta entre DOCUMENTACIÓN Y TERMINOLOGÍA

- Ambas se ocupan de estudiar textos en lenguajes especiales.
- La conexión más directa entre ambas radica en lo siguiente:

La documentación se ocupa sobre todo de documentos de diferentes tipos, que son objeto de tratamiento científico, y la Terminología es fundamental en documentación para:

- Ordenación de conceptos u ordenación de conocimientos científicos mediante la creación de sistemas de conceptos (Clasificaciones).
- Transferir conocimientos a través de la enseñanza y la formación de nuestros especialistas.
- Para comunicar conocimientos de un lenguaje a otro evidentemente a través de unos signos lingüísticos.
- También nos sirve para resumir y sintetizar la información de tipo científico.
- Almacenar y recuperar información mediante los Lenguajes Documentales es decir de la Indización.

Está claro que se relaciona con el Análisis Documental tanto de contenido (interno) como formal (externo).

Nos atañe principalmente en el contenido.

Formula la terminología a través de los descriptores, no sólo proporciona el término del descriptor sino que a cada uno le da un peso específico. De tal manera, que al conocer ese peso específico podrá indizar utilizando los descriptores con mayor o menor peso científico según la especificidad que yo le quiera dar, que además me supone poder recuperar con mayor o menor eficacia.

3.2.– MÉTODOS DE TRABAJO EN TERMINOLOGÍA

La mayoría de los diccionarios son meras combinaciones de términos, pero no elaboran una conexión entre los términos.

En cambio, en documentación, los tesauros exigen asociación entre sus términos, es decir, hay que analizar no sólo los términos sino también los conceptos. La norma ISO 919 es una guía para la elaboración de tesauros sistemáticos, de vocabularios sistemáticos.

Antes de comenzar con la labor terminológica hay que seguir una serie de pasos:

- Se debe precisar el campo de estudio, la disciplina o subdisciplina con la que vamos a trabajar. También debemos saber el número de conceptos con los que vamos a trabajar (VOLMENT).
- Definir el idioma con el que trabajaremos
- Hay que realizar una ficha con el dato terminológico que vamos a incluir.

Existen diferentes tipos de bases de datos terminológicas, que se pueden dividir en dos grandes grupos:

1.– Bases de datos de diccionarios:

Orientadas sobre todo a la traducción. Es una base de datos como un diccionario impreso. La utilidad es la misma. La ventaja es que tiene una actualización muy rápida. Por lo tanto, es una herramienta muy útil a la hora de indizar. Son las que más se utilizan para la traducción automatizada, sin embargo, ésta no está dando los resultados esperados.

2.– Bases de datos vocabularios:

Estos ya tiene un concepto mucho más útil, es un concepto de tesoro. Los registros que recoge no sólo son los conceptos, sino la relación de esos conceptos con otros conceptos similares. Es decir, actúa como un tesoro. Pueden ser monolingües o multilingües

Las principales bases de datos mundiales son:

- La base de datos TERMIUN– de Canadá: sus lenguas oficiales son el Inglés y en francés.
- NORMATERM, en Francia, esta base de datos la lleva a cabo AFNOR
- EURODICAUTON (CEE)
- TERMINOLOGIA, Alemana.
- TERMOC , Noruega
- TERMCAT (Cataluña)
- UZEI (Vasca)
- HISPANOTERM (Madrid), esta comenzada por Criado del Val, pero sin embargo todavía no está acabada.

La norma ISO 5127 – 2 de 1983 que regula la definición exacta y sus equivalencia en Francés e Inglés de términos en Documentación.

Son términos extraídos de glosarios, diccionarios, también son documentos extraídos de la IFLA.

La función más importante que tiene ya no es sólo ayudar a documentalistas sino ayudar en la transmisión del conocimiento de un país a otro.

29 / 11 / 99

TEMA 4.-

4.1.- LENGUAJES DE INDIZACIÓN

Esquema sobre el concepto de indización. Este consta de 3 elementos:

1.- Representación del contenido

a.- Lenguaje especializado

! características:

- vocabulario especializado
- giros especializados
- rasgos morfosintácticos propios
- verbos, adjetivos, y otras categorías distintas de las sustantivas.

b.- Problemática de la representación

Triángulo semiótico

Problemas del indizador

- relación término-concepto
- relación concepto- concepto

2.- Formas de representación

a.- Sistemática

- clasificaciones

b.- Alfabética

- Por extracción ! palabras clave
- Indización libre ! descriptores libres
- Por asignación ! descriptores

encabezamientos de materia

3.- Recuperación

El concepto de indización consta de tres elementos como hemos visto anteriormente:

- Representación del conocimiento
- Formas de representación
- Recuperación

En documentación, el documento es la materia prima con la que se van a realizar todas las tareas de documentación.

Al hablar de documento, nos referimos a documento científico.

¿ Definición de indización:

La palabra indización es muy reciente. Aparece por primera vez en el año 84. La RAE la incluye con dos acepciones:

- 1.- Tarea de realizar índices
- 2.- La indicación es la forma de registrar ordenadamente datos e informaciones para elaborar índice.

¿Qué se entiende en documentación por INDIZACIÓN?

La operación que consiste en representar el contenido de un documento mediante un conjunto de términos que especifican los temas de que trata ese documento. Con el fin de ayudar a su posterior recuperación de la información. Actualmente, se le está añadiendo a esta definición la coletilla de recuperación idónea de los documentos, es decir, no consiste en recuperar cuantitativamente, sino cuantitativa y cualitativamente.

Esta definición consta de una serie de elementos:

- Representación del conocimiento
- Formas de representación
- Recuperación

1.- Representación del contenido:

La representación del contenido exige una tarea previa:

- La comprensión del documentos, saber el criterio de ese documento. Esto será independiente del sistema de indización que utilicemos, de ahí que los profesionales conozcan las disciplinas que indizan en un grado bastante elevado. Esto topa con un problema siempre, la imprecisión del lenguaje, la imprecisión de los signos lingüísticos, ya que a veces no se corresponde con el lenguaje natural. Por ejemplo: El lenguaje especializado no siempre corresponde con el lenguaje natural.
- El indizador debe comprender la información que tiene en ese documento. El documentalista tiene que conocer gran parte de las disciplinas que tiene que indizar.

1. A.- Lenguaje especializado

La norma DIN 2342 dice lo siguiente con respecto al Lenguaje Especializado: Área de la lengua que aspira a una comunicación unívoca y libre de contradicciones en un área especializada concreta. Cuyo funcionamiento se basa fundamentalmente en la terminología establecida. Es una lengua propia para cada una de las áreas del conocimiento. Existen tantos lenguajes especializados como áreas.

No hay un único lenguaje especializado, hay muchos, dependen del área.

La indización (su objeto) son esos lenguajes especializados, dentro de estos tenemos cuatro características:

! Los lenguajes especializados disponen de un **vocabulario especializado**.

Smith habla de tres tipos de vocabularios especializados, que son: términos, cuasitérminos, y jergalismos.

• Los términos:

Aquellas palabras técnicas cuyo contenido está fijado por definiciones concretas, se caracteriza por su univocidad.

Palabras técnicas cuyo contenido está perfectamente definido.

Ej: biblioteca.

• Cuasitérmino

Son también palabras técnicas (tecnicismos) pero no definidos, su definición no está definida con antelación.

Los cuasitérminos sirven para definir un concepto y normalmente nos llevan a equivocación.

Si se utilizan los cuasitérminos es porque nos lo han impuesto, vienen impuesto por la práctica su uso. El significado varía según la disciplina en la que nos movamos.

Ej: Red: Red de transmisión de datos

Red de pescador, red para pescar

Otra característica que diferencia los términos de los cuasitérminos es que los términos están aceptados universalmente, su definición tiene un valor único. Mientras que los cuasitérminos tiene problemas e comprensión, plantean dudas.

Las reglas de indización articularán fórmulas para definir estos cuasitérminos.

• Jergalismos

No pretenden ni la exactitud, ni la univocidad de los anteriores, son expresiones del lenguaje norma.

! Giros especializados:

Los giros especializados dan lugar a la fraseología.

En los Lenguajes Especializados a parte de palabras especializadas hay giros especializados, es decir,

combinaciones sintácticas de al menos 2 elementos de un léxico especializado, son frases, por eso da lugar a la fraseología: que es el conjunto de giros especializados.

! Rasgos morfosintácticos propios

HOFFMANN los define con las siguientes características:

- El verbo pierde sus características temporales y la mayoría de las veces aparece en presente y en tercera persona del singular.
- A menudo el verbo suele estar en pasiva y como característica general el verbo tiene poca importancia o un papel poco relevante. Es el sustantivo el que mayor importancia léxica tiene.

Se suele emplear el sustantivo con prioridad del singular al plural.

- Y el adjetivo aparece con bastante poca frecuencia.

No obstante, algunos autores dicen que el verbo y el adjetivo pueden tener un papel importante (dado que se expresan acciones, movimientos. Localizaciones, etc...) Aunque hoy por hoy lo que se utiliza es lo dicho anteriormente.

1.B.– Problemática de la representación

Se supone que al tener un lenguaje especializado, se deberían tener menos problemas, pero esto no es así. La terminología demuestra que el hombre no sólo utiliza un término para expresar sólo un concepto, sino que utiliza más. De ahí viene la problemática. Por lo tanto de aquí surgen los problemas a la hora de representar los contenidos, este problema se ha representado con el :

! Triángulo semiótico, que se representa como:

objeto: Representa la realidad

Término: representa la expresión elegida

Concepto: Pensamiento referido a ese concepto

Fue SOERGEL el primero que apunta este esquema a las ciencias de la documentación. Dice que en este esquema:

* El objeto: es la existencia real, efectiva y objetiva de una cosa que puede ser concreta o abstracta.

* Que el concepto: unidad de pensamiento constituido por aquellas características que se atribuyen a un objetivo o a una clase de objetos.

Existen dos tipos de conceptos:

1.– Individuales: representan un sólo objeto

2.– Generales: representan a un conjunto de objetos, que deberían converger en una serie de características comunes.

* Y que el término es el mínimo de palabras que designan un concepto del lenguaje especializado. A efectos de

aplicación para términos de indización hay dos clases de términos:

1.– Expresiones léxicas:

Es una sola palabra que es comúnmente admitida.

Ej: Archivo: Por supuesto se refiere a un concepto unívoco.

2.– Expresiones no léxicas:

Formadas por más de una palabra que designan un concepto a una realidad.

Ej: Centro de documentación.

! Problemas del indizador a la hora de indizar

– relación término – concepto

* sinonimia,

* polisemia

* homonimia

– relación concepto – concepto

Sobre todo la relación entre el termino y el concepto es importantísima, porque nos da las características de la indización. Si la ambigüedad no se controla los documentos recuperados no van a satisfacer las demandas de los usuarios.

* La relación TÉRMINO – CONCEPTO

Un concepto puede designar varios términos, y varios términos pueden designar un concepto. SOERGEL es uno de los máximos representantes a la hora de estudiar esto. Él dice que ha de existir una relación unívoca entre los dos, pero suelen aparecer problemas de sinonimia, polisemia y homonimia.

Conceptos términos

SINONIMIA	A	X
		Y
		Z
POLISEMIA	A	X
	B	
	C	
HOMONIMIA	A	X
	B	X

	C	X
--	---	---

* Sinonimia: (Un concepto con dos o más denominaciones intercambiables arbitrariamente). Ej: Coche, automóvil.

* Polisemia: (Reunión de varios significados en una misma palabra, suelen estar relacionados). Ej: pie como tal, pie como medida de longitud, pie como base de algo).

* Homonimia: (varios términos se expresan de la misma manera, pero los conceptos a los que se refieren no tienen nada que ver entre sí, no tienen relación).

Para distinguir la homonimia de la polisemia:

¿Cuándo un término es polisémico o hay homonimia?

1.- Conciencia lingüística de los hablantes:

Ej: Si hablamos de un bombero estaremos hablando de la llama de fuego

2.- Relación existente entre las palabras:

Pie es polisémico ya que los tres tienen el mismo concepto

3.- Distinción entre la propia etimología de las palabras:

– La polisemia será: aquellos términos cuyo origen etimológico es el mismo. Si yo hablo de llama : del latín hama

de fuego

de animal

Existen también tres tipos de homonimia:

1.- Homógrafos: son los que se escriben igual, tiene coincidencia gráfica vino (beber), y vino de verbo.

2.- Homófono: se escuchan igual pero no se pronuncian igual – Haya, halla.

3.- Homónimo pleno: Son sonidos y sus gráficas es la misma (Banco de escribir y banco de sentarse).

* Problemática respecto a la relación concepto – concepto

En la relación concepto – concepto existen cuasisinónimos (conceptos que son similares a los sinónimos pero son distintos en alguna manera)

Automóvil y vehículo.

13 / 12 / 99

2.- Formas de representación del contenido

Con respecto a las formas de representación del contenido, la definición hace referencia a cualquier tipo de indización.

Hay que distinguir entre indización y clasificación.

La distinción entre clasificación e indización es una distinción formal:

Indización: operación que consiste en escoger los términos más apropiados para representar el contenido de un documento.

Clasificación: es una operación que describe el contenido de un documento determinado, su tema principal y a veces uno o dos temas secundarios que traducen a términos más apropiados que figuran en el lenguaje documental que se utiliza.

INDIZACIÓN

Las primeras representaciones de los documentos han sido formales, han pasado a un segundo plano, hoy en día es el contenido de los documentos.

El análisis documental del contenido es lo que más se toma en cuenta y a lo que más importancia dan los autores, ya que agrupamos a la clasificación junto con la indización. Se puede decir que hay dos formas de indización:

- a.- Indización sistemática ----- clasificaciones
- b.- Indización alfabética -----
 - 1.- por extracción ; palabras clave
 - 2.- indización libre ; descriptores libres
 - 3.- Por asignación; descriptores y los encabezamientos de materia.

Por supuesto, según el tipo de indización que se utilice, cada uno exige un tipo de lenguaje diferente.

– La indización de tesauros exige un control estricto de indización.

La finalidad de la indización es la RECUPERACIÓN.

3.- La recuperación

La recuperación está también relacionada con los distintos sistemas de indización.

También está estrechamente relacionada la recuperación con la política del centro (si son más especializados, menos, etc...).

También es relevante el sistema de recuperación del centro.

FASES DE INDIZACIÓN

De los autores más relevantes se han resumido los conceptos más importantes:

- 1.- Según VAN SLYPE, las fases de la indización constan de cuatro etapas:

- a.- Hay que revisar el contenido de los documentos
- b.- seleccionar los conceptos
- c.- se traducen los conceptos en descriptores
- d.- Se establecen enlaces sintácticos entre descriptores

Otros autores engloban dos etapas de Van Slype en una, así, tenemos el caso de:

2.- FUGMAN, que habla de:

- a.- El reconocimiento y selección del contenido de los documentos
- b.- Representación del contenido de los documentos

3.- LANCASTER:

- a.- Análisis conceptual
- b.- Traducción a un lenguaje de indización.

4.- La norma ISO – 5963 habla de 3 etapas:

- a.- Examen del documento y definición de su contenido
- b.- identificación y selección de las nociones principales del contenido
- c.- Selección de los términos de indización.

* Aunque se diga que son 2, 3 o 4 las etapas en las que se engloban, se engloba todo en la misma, ya que se piden los mismos criterios.

Conclusiones:

- Hay que unificar la terminología a emplear a la hora de plasmar los contenidos (por eso es necesario unificar la terminología)
- Es difícil establecer fases rígidas puesto que no todos los autores están de acuerdo con las mismas fases. Es difícil además puesto que hay diferentes sistemas de indización, depende además de la mayor o menor exhaustividad con la que queremos indizar el documento.
- Es bastante necesario establecer unas fases generales comunes a cualquier proceso de indización, pero luego, cada uno tendrá unas etapas específicas.

Estas fases generales deben responder al qué y al cómo (cómo voy a transmitir esa información) . De como yo responda se derivarán las fases generales de indización

Análisis del documentos y representación del documento.

Con lo visto hasta ahora podemos decir que las fases de indización serán dos:

1.- Análisis del documento

2.- Representación del documento.

A cada fase le corresponden unas etapas.

FASES	ETAPAS				
	Palabras clave	Descriptores libre	descriptores (tesauros)	Materias	Clasificaciones
Análisis del documento	Si indizo con palabras clave las etapas serán:	Análisis	Análisis	Análisis	Análisis
		Identificación	Identificación	Identificación	Identificación
	Identificación y selección	y selección	y selección	y selección	y selección
Representación	Extracción	Extracción	Descriptores controlados	Traducción	Traducción
		Asignación			
		Control			

– Índice por cualquier forma menos por palabras clave hay tres etapas:

En identificación y selección en palabras clave son de palabras (extraer palabras) es en definitiva identificar y seleccionar palabras.

Mientras que en análisis, identificación y selección son de contenido, se extraen conceptos del documento, en definitiva se analizan, identifican y seleccionan conceptos

Extraiga el concepto que extraiga tengo que traducirlo del lenguaje libre al controlado.

Explicación:

Respecto a la primera fase, todos los autores están de acuerdo, en lo que ya no están tanto de acuerdo es en las normas como en el método preciso para trabajar en estas etapas.

No hay una metodología detallada para saber como analizar, identificar y seleccionar.

Últimamente, la literatura anglosajona trabaja sobre ello, y habla de seguir unas pautas:

1ª FASE :ANÁLISIS DEL DOCUMENTO

1ª ETAPA : ANÁLISIS DEL CONTENIDO,

1.- Lo primero que hay que hacer es realizar una primera tomo de contacto con el documento, ver de que trata. Para ello, se debe leer, aunque esto es realmente imposible para el indizador, por lo tanto, el indizador lo que hace es que tiene unas fuentes para nutrirse, para saber de lo que habla el documento impreso sin necesidad de leerlo entero. Observa pues el:

– título, este deberá explicar el contenido del documento, pero como este es muy objetivo por parte del autor, podemos decir que no es una fuente muy segura.

– Pasamos al resumen, que es bastante fiable, ya que lo que hace es describir el contenido de forma sintética. Pero, depende de quién halla hecho el resumen, si lo ha hecho el autor se puede distorsionar la información, si

lo hace el indizador, corre el peligro de no captar exactamente lo que quiere decir el autor.

- Una forma es obligar al autor a incluir en sus libros un resumen de su publicación, además se le exige también que ponga una serie de palabras clave. Así, el indizador con ese resumen y las palabras clave (que son de buena fuente) puede saber cual es contenido de ese texto.
- Las revistas periódicas incluyen el resumen, aunque otros textos no, esto es un problema para el indizador.
- También son los sumarios una buena información, estos aparecen en publicaciones periódicas.
- Las introducciones de los documentos, aunque corren el mismo peligro que el título, suelen ser a veces subjetivas.
- El propio cuerpo del documento, tablas, títulos del capítulo, conclusiones generales, etc...
- cuantas más utilice el autor, mayor fiabilidad a la hora de la indización.

En la segunda etapa: **LA IDENTIFICACIÓN DEL CONCEPTO**, consiste en reconocer los conceptos que representan el contenido de los documentos.

Uno de los problemas es la necesidad de normalizar el método de identificación y selección.

La norma ISO 5963 habla de : Métodos para el análisis de documentos, las determinación de su contenido. Esta norma se limita a dar unas recomendaciones generales para identificar conceptos.

Se dice: el indizador debe crear el mismo una plantilla de identificación, es decir, recomienda hacer una especie de esquema previo con una serie de categorías o conceptos ya plasmados que puedan identificar los conceptos de ese nuevo documento.

La plantilla debe ser de dos formas:

- a.– O bien en facetas
- b.– O bien por temas

¿Para qué hacer la plantilla?, para proporcionar unas pautas de orientación al indizador, para unificar unos criterios, para saber que clase de conceptos debemos incluir en la clasificación, y para facilitar la recuperación. También nos puede ayudar para elaborar un lenguaje de indización.

Esta plantilla para su elaboración necesita seguir las siguientes pautas:

- Este esquema se debe adaptar a la materia que vamos a indizar
- La plantilla debe estar adaptada al tipo de documento
- Para hacerla o realizarla nos podemos ayudar de clasificaciones ya existentes.
- Conviene también hacer una primera indización rápida de unos pocos documentos y extraer unos conceptos claves.

Esta **plantilla puede ser a facetas o materia**. Las **plantillas realizadas a facetas** (el que lo estudio fue Van Slype) dice que de todos los documentos él solo ha detectado 8 facetas:

- Fenómenos: acciones naturales que escapan a la mano del hombre
- Procesos: acciones provocadas por los hombres.

- **Materiales:** elementos materiales, naturales o no, sobre los que actúan los fenómenos y los procesos.
- **Organizaciones:** entidades complejas creadas por seres vivos
- **Seres vivos:** microorganismos, vegetales, animales, seres humanos, etc...
- **Equipamientos:** medios contruidos por el hombre para operar sobre los materiales o los fenómenos a través de los procesos.
- **Propiedades:** Características ligadas a uno o más fenómenos, procesos, materiales, seres vivos, o equipamientos.
- **Disciplinas:** Distintas ramas del conocimiento

Plantillas realizadas por temas: cuya característica es extraer los temas principales del documento.

Ejemplo; Documentos sobre el insecto bibliofago en algunos documentos:

Plantilla por facetas – Materiales – papel – papiro – Seres vivos – animales – personas – Proceso – restauración Extraigo además conceptos más secundarios que los acoplo en los grandes temas que tenía anteriormente.	Plantilla por temas – Materiales – Seres vivos – Proceso Es simplemente hacer la plantilla en base a grandes temas, extraigo sólo del documento los conceptos principales.
---	---

20 / 12 / 99

Tercera etapa del análisis: SELECCIÓN DE TÉRMINOS

El indizador no tiene que leerse el documento entero. Está permitido que el documentalista no se lea el documento entero para extraer la selección de los términos, está se ve influida por tres factores:

- por la profundidad de la indización
- por el tipo de documento y su contenido
- por los propios intereses de los usuarios.

A.– Por la profundidad de la indización

* **Según la indización** se haga de una forma más general o más específica así seleccionaré los términos , más generales o más específicos.

* **El grado de especialización del centro.** Según sea más o menos especializado se seleccionarán los términos más o menos especializados.

* **El sistema de indización** también influye a la hora de seleccionar los términos.

Tipos de indización: por medio de descriptores libres, palabras clave, encabezamientos de materia.

* **Según como indice el centro.** Es sistema de indización influye a la hora de seleccionar los términos. No es lo mismo indizar con encabezamientos de materia que con palabras clave.

El sistema que permite mayor profundidad son los siguientes en grado de mayor a menor: descriptores libres y palabras clave, y descriptores controlados (los que se emplean en los tesauros), los encabezamientos de materia y clasificaciones.

Estos permiten representar los conceptos

B.– Por el tipo de documento y su contenido

No es lo mismo tener una monografía o una publicación periódica, el tipo de documento proporciona una visión de contenido distinta. Por lo tanto, la información de un documento no depende sólo de la extensión de ese documento. A la hora de seleccionar los términos hay que tener en cuenta qué tipo de documento es, es decir si es monografía, publicación periódica, etc...

Van Slype da unas pautas a la hora de elegir un documento, las nociones a la hora de determinar el tema de indización.

Pautas para extraer las nociones del documento:

existen cosas en un documento que el indizador debe retener y otras que no.

El indizador debe retener;

- Nociones explicativas del documento, información explícita, que el indizador comprenda que va a responder a las demandas de los usuarios.

- También debe retener el mensaje implícito del documento. Aquello que el autor no dice explícitamente, pero que a través de su lectura se entiende como tal.

Cosas que el indizador no debe retener:

- Aquellas nociones presentes en el documento pero que el autor dice que no va a tratar.

- Nociones que figuran sólo a título de ejemplo

- Todas aquellas nociones que aunque están presentes en el documento no sean tratadas con la suficiente profundidad para que le sea interesante al usuario.

C.– Por los propios intereses de los usuarios

¿Se plantea en cuestión de selección los términos en base al usuario para el que está pensado el centro, o, seleccionar los términos para otros usuarios que también los demanden?

Algunos autores precisan que sólo debe ser seleccionado el término de acuerdo con el usuario real. Aunque también es importante seleccionar el término de acuerdo al usuario real y potencial.

Influye en la selección de los términos el interés del usuario. Algunos autores dicen que sólo se debe atender a las necesidades de los usuarios de ese centro de documentación. Otros dicen que no se sea tan estricto a la hora de seleccionar el vocabulario para un usuario concreto, sino que lo entiendan los demás usuarios, es decir, favorecer que otro tipo de usuarios lo puedan utilizar.

Pero con el uso de redes se está popularizando el acceso a las bases de datos con un lenguaje más accesible, ya que cada vez tiene acceso más público y no sólo pensando en el usuario real sino también en el usuario potencial.

2ª Fase: Representación del contenido:

Esta fase es más problemática que la primera (ANÁLISIS) ya que no tiene muy bien definida las etapas.

Aquí la representación del contenido depende del sistema de indización que elijamos. Depende si es por modo de palabras clave, descriptores libres, encabezamientos de materia, etc...

Sistema	Procedimiento
1.- Palabras clave	Extracción
2.- Descriptores libre	Control / extracción
3.- Descriptores controlados por materias; clasificaciones	Traducción

tema 6.-

SISTEMAS DE INDIZACIÓN

La indización puede ser por:

6.1.- Sistema de indización por extracción

También se llama indización por entrada.

Consiste en tomar de los textos originales los términos con significación real e introducirlos en el sistema de búsqueda. Se realiza a través de **palabras-clave** ! son palabras significativas o no vacías extraídas del título, del resumen del texto de los documentos, que se usan para representar su contenido sin someterse a ningún tipo de control terminológico. Es decir, palabras que extraigo tal cual del documento, como el autor lo dice.

Palabras vacías ! son aquellas que no tienen ningún contenido. Artículos, preposiciones, pronombres, numerales, interjecciones, en cambio, nombres, adjetivos y adverbios son palabras no vacías.

La extracción de palabras claves, se produce en la primera fase: la del análisis.

METODOLOGÍA DE INDIZACIÓN POR PALABRAS CLAVE

Lo normal es que las palabras claves las extraiga el ordenador en base a la lectura de ese documento.

Lo primero que hay que hacer es introducir un listado de palabras vacías (también llamadas antidicionario), también se almacena en este antidicionario todos aquellos caracteres que sirven de puntuación (puntos,

comas, comillas, interrogaciones, etc...)

Una vez construido el antídiconario, el segundo paso es ya con los documentos, indizar, todas las palabras significativas de ese documento las compara con ese antídiconario, y el ordenador selecciona las palabras que no aparecen en el antídiconario.

Con las palabras seleccionadas se construye una lista y voy formando las palabras clave, puede añadir tantas como quiera.

El ordenador sabe con qué frecuencia aparecen las palabras, si aparece una nueva la introduce, si aparece una que ya está anota su frecuencia. Con ello, obtengo todas las variantes que parten de una misma raíz. Y además, obtengo una base de datos con todas las palabras que tengan una raíz común.

Este sistema de EXTRACCIÓN O DE INDIZACIÓN POR PALABRAS CLAVE, realmente cuando ha tenido su utilidad es a través del uso del troncal, truncamiento, sino no se podría obtener una optima utilización.

Para recuperar utilizo el TRUNCADO y recupero por palabras clave.

Ventajas de esta indización

- 1.- rapidez, y bastante económico al no exigir ningún tipo de control
- 2.- La palabra clave elegida va a representar con mucha exactitud lo que quiere decir el documento. Se trata por lo tanto de una indización muy específica, también es muy exhaustiva ya que puede indizar con todo tipo de acepciones de una misma palabra. (*)

Puede clasificar todas las palabras que quiera con una misma raíz.

- 3.- Con el seguimiento de las palabras clave de un sistema se aprecia exactamente la evolución de las palabras (si antes se utilizaba política social, y ahora se utiliza sociedad política) (*)

Con el seguimiento de una palabra clave de un sistema se aprecia el seguimiento del lenguaje de una materia. Se observa la evolución del lenguaje en una materia determinada, por eso es más vivo.

Inconvenientes

- 1.- No tienen ningún control sobre los polisémicos y sobre los sinónimos
- 2.- Falsas combinaciones (centro de documentación). No busca esto, ya que la de es una palabra vacía, por lo tanto el sistema sólo buscar todo aquello que tenga que ver con centros y todo aquello que tenga que ver con documentación, no encuentra por lo tanto centros de documentación.

Actualmente, con las nuevas técnicas, el documentalista ya incorpora esto, y el sistema permite alguna palabra vacía dentro de las palabras claves.

- 3.- Espacio:

Existe un problema de almacenamiento ya que son listas enormes, no tiene delimitación, esto genera también problemas.

- 4.- Sólo contempla conceptos contemplados explícitamente en un documento, por lo que conceptos implícitos

no aparecen.

5.- Los homógrafos de palabras vacías no se almacenan

(sobre como preposición, o sobre como sustantivo de carta, por lo que ninguna de las dos son contempladas).

6.- Tampoco almacena las siglas seguidas de punto.

O.N.U – no lo reconoce.

Como las palabras las extrae del documento no tiene en cuenta errores tipográficos y mecanográficos. Esta palabra no las recupera ya que estaba mal escrita.

Problemas en la recuperación:

1.- No controla las distintas formas gramaticales, por lo tanto, los singulares y plurales, es frecuente que no me las controle. Todos los supuestos problemáticos como sinónimos, etc... es el indizador quien los tiene que controlar.

6.2.- La indización libre:

Es un tipo de indización intermedia, ya que a veces que utiliza palabras clave y descriptores libres.

En esta indización no se usa ningún vocabulario determinado, es decir, el indizador tiene libertad absoluta para extraer términos directamente del documento, pero también tiene capacidad de asignar él términos que no aparecen explícitamente pero si pueden aparecer implícitamente en el documento.

La forma de los términos si que debe estar controlada, que atenderá a las normas del centro de documentación. Si el centro dice por ejemplo que no hay que utilizar femeninos y sólo se podrá utilizar masculinos, la norma debe aplicarse al descriptor libre en ese sitio.

También tiene un cierto control semántico, se controla que no proliferen sinónimos, polisémicos, etc...

La normalización del centro se aplica al descriptor libre pero no a la palabra clave.

Hay un control semántico para suprimir los sinónimos.

No tiene límite el engrosar la lista de descriptores libres.

Definición de descriptor libre:

Términos de indización que representan conceptos destacados por un proceso intelectual a partir de los documentos almacenados en un sistema documental dado.

Diferencia con respecto a las palabras clave: es que exigen un cierto control intelectual.

El descriptor libre puede emplear frases (palabras compuestas) por ejemplo: Centro de Documentación.

Se someten a un cierto control de forma, el ordenador genera un listado de descriptores libre, pero en este listado si hay un cierto control, las equivalencias se controlan. Ejemplo: en el listado un término equivale a otro, por lo tanto la polisemia y la sinonimia se ven mitigadas.

Cuando yo indizo por descriptores libres paso por una serie de etapas:

- 1.– Establecer unas reglas básicas para controlar las formas, estas reglas son establecidas por el centro de documentación a fin de controlar las formas de los términos. Se hace corrección de errores. Los listados son revisados por el indizador, los reenvíos, errores tipográficos, etc...
- 2.– Editar una lista definitiva que sirva de guía.

10 / 01 /2000

Método de indización por descriptores libres:

El método de indización por descriptores libres pasa por una serie de etapas:

- El establecimiento de unas reglas básicas del control de la forma de esos descriptores libres.
- Los documentos se han de indizar siguiendo las pautas establecidas previamente
- Analizar la lista de los descriptores utilizados en la segunda etapa para corregir errores.
- Hacer un semicontrol semántico
- Edición de la lista definitiva de esos descriptores.

1.– El establecimiento de unas reglas básicas del control de la forma de esos descriptores libres:

Consiste en: los descriptores libres se pueden hacer de dos formas:

- o bien extrayéndolos directamente del documento
- o bien pueden haber sido asignados por el propio indizador (esta es una diferencia con respecto a las palabras clave)

Hay que realizar un control morfológico. El control morfológico no está sometido a ninguna regla. Las reglas las establece el propio centro indizador. Sí hay directrices, pero no reglas.

En cada país se observan unas determinadas tendencias a la hora de indizar con descriptores libre.

Es aconsejable seguir los criterios de cada país.

En el caso de España, estas reglas, las más conocidas y las que se aconseja seguir son;

- Norma UNE 50–106–90
- Encabezamientos de materia y reglas de catalogación
- Normas de la Universidad de Sevilla, son una lista de Encabezamientos de materia que también se utilizan para indizar en diversos centros.

Aspectos básicos a regular en los descriptores:

- Lengua
- Uso de formas nominales
- Género y número.

Lengua:

Los descriptores deben utilizarse con respecto a la lengua del centro de indización, esta regla se aplica también a los nombres propios.

Ejemplo: Debemos utilizar en España Londres y no LONDON

Sólo adoptaremos palabras extranjeras cuando no exista un equivalente en español. O sólo cuando la palabra utilizada esté tan extendida que no lleve a equivocación. Ej: Marketing.

Uso de formas nominales:

En cuanto al uso de formas nominales se debe utilizar sólo formas nominales y no verbos ni adjetivos, en todo caso sustantivizar. Nunca utilizar indizar, sino indización.

Género y número:

En cuanto al género se usa el masculino al ser más genérico. Sólo se utiliza el femenino si se hace referencia concreta a algo explícito de la mujer.

En cuanto al número es uno de los aspectos en los que más diferencias se encuentra en los países.

Anglosajón: se utiliza el singular en lugar del plural, dependerá también del tipo de nombre.

Franceses y alemanes: utilizan casi siempre el singular.

Nosotros en cuanto al género hacemos dos distinciones:

– Si son entes concretos o puede haber sustantivos cuantificables (Objetos susceptibles de ser cuantificados)

Se ha de responder a la pregunta ¿Cuántos?

Entonces se ponen en plural

Si el sustantivo es no cuantificables ¿Cuanto? ¿Cuantas? Son siempre en singular: Ej: Información.

– Hay excepción con partes del cuerpo, se expresa en plural si el órgano es más de uno, si el órgano es uno sólo se expresa en singular.

Con respecto a los conceptos abstractos, son propiedades, sistemas de creencias, actividades y disciplinas, la regla nos dice que son en singular.

2.– Indización de documentos con los descriptores a los que ya hemos aplicado esas reglas

Aquí hay que aplicar el mismo descriptor al mismo concepto.

3.– Listas de descriptores. La corrección de errores

Una vez aplicado esos descriptores, se elabora una lista provisional con esos descriptores, en ella se pone:

– el descriptor

- y el número de frecuencias con las que aparece en esos documentos indizados (número de registros en los que ha aparecido)

Esta hoja sirve para detectar errores. El tipo de errores que se detectan son: errores ortográficos (si existen siglas con o sin puntos), faltas de mecanografía, singulares y plurales, y también si existen conceptos sinónimos (centro de documentación, centro de orientación).

4.- Control semántico

A través de la lista provisional se hace un semicontrol semántico, intentando poner equivalencias, se puede también establecer jerarquías e equivalencias.

5.- Publicación definitiva del listado de descriptores

Esta es la que empleamos para indizar.

Ventajas e inconvenientes de estas listas:

Ventajas:

- es bastante específica, permite identificar conceptos que están implícita y explícitamente.
- Puede utilizar expresiones léxicas y no léxicas.
- Es bastante sencillo indizar de esta forma, dado que utiliza pocas reglas.
- Es un tipo de indización más apto para ciertas disciplinas, por ejemplo:: ciencias que utilizan un lenguaje específico como ciencias experimentales. No es nada recomendable para ciencias sociales y humanidades.

Inconvenientes:

- Grado de imprecisión elevado, el indizador no tiene un lenguaje controlado, consecuencia, una recuperación de la información con baja calidad.
- Lo que mejor resume a este sistema de indización es la Ambigüedad.

6.3.- INDIZACIÓN POR ASIGNACIÓN

Características generales:

- Dada la imprecisión de nuestro lenguaje, en esta forma de indización hay que elegir una expresión y sólo una.
- Exige un control semántico bastante fuerte.
- Se pueden utilizar más de un término (términos compuestos), lo cual elimina la ambigüedad respondiendo mejor a preguntas específicas que hace el usuario.
- Se parte de un vocabulario dado, no admite otros términos de indización que no sean los ya establecidos.
- Se produce un control.

- Exige rigidez puesto que es complicado el introducir términos nuevos.
- La forma de indización: utiliza ENCABEZAMIENTOS DE MATERIA (palabras simples o compuestas) y los DESCRIPTORES CONTROLADOS.

Con respecto a los descriptores se sigue una serie de normas:

- La indización a través de los descriptores controlados surge en los años 60. Lo inventó CALVIN. Este tipo de descriptores están basados en conceptos globales sin importar el número de palabras que incluye el concepto. Pueden ser de la extensión que se quiera.

Hay tres tipos de descriptores:

- 1.– Atendiendo a su información
- 2.– Atendiendo a su contenido
- 3.– Atendiendo a su estructura

- 1.– Atendiendo a su información:

Pueden ser:

- a.– Primarios: cuando se refieren al tema principal del documento
- b.– Secundarios o auxiliares: son descriptores que no tiene la carga informativa del descriptor primario *pero añaden o restan información al primario*, y en este sentido añaden información al primero.

Pueden ser también descriptores que *se refieren a un tema secundario* del que trata el tema.

Esto va a depender del centro de documentación

- c.– Infraconceptos: Partículas que por ellas mismas aisladas no significan nada, pero que cuando acompañan al descriptor les da un significado.

Ej: micro, maxi, etc....

- 2.– Atendiendo a su contenido:

- a.– Cronológicos: (fechas, años, décadas, los siglos se suelen poner en números romanos).

- B.– Onomásticos: o también llamados identificador/es, se refieren a personas citadas en el texto, es decir a nombres propios o instituciones.

Los onomásticos son complicados, la forma es que se pongan con la traducción del idioma del centro de documentación. El problema es cuando la traducción se realiza con nombres que no son latinos.

Si viene la profesión de la persona citada, la norma dice que como identificador se tiene que poner el nombre y entre paréntesis su profesión.

- C.– Geográficos: estos son muy complejos porque es muy cambiante dado que hay nombres que no se conservan, ej: Unión Soviética.

La nomenclatura, se utiliza la del propio país, ahora en España surge un problema con este tipo de descriptores con respecto a las lenguas vernáculas. Ej: que se pone San Sebastián o Donostia.

Hay que poner también los descriptores conservando su jerarquía. País, provincia, pueblo, etc...

También hay que tener cuidado con ambigüedades y polisemia.

Ej: Latino América, Iberoamérica, esto hay que especificarlo.

3.- Atendiendo a su estructura:

a.- Uniterminos

b.- Sintagmáticos, que son más de un término

Condiciones que debe reunir un descriptor

1.- Nunca verbos

2.- Como pueden ser sintagmáticos, por lo menos una de las palabras debe ser un sustantivo.

3.- Debe representar un sólo concepto.

4.- Los sintagmáticos no deben emplear conjunción.

Método para construir estos descriptores:

Puedo construir descriptores de dos maneras:

1- Por medio de una unión **morfológica**, con 1, 2 o más descriptores (sin que varíe el significado de ellos)

Ej: Política – económica

2.- **Unión semántica**, unión de 1,2, o más palabras, donde transformo su significante.

Ej: si una documentación a distancia, obtengo: teledocumentación.

Normas para la realización de descriptores:

* si es unitérmino = un sustantivo

* si es una frase = será una frase sustantivada

- Si hay opción, se prefiere el singular para los no cuantificables.
- Deben adoptar los descriptores una forma desarrollada, es decir, no se deben admitir las siglas salvo que éstas estén totalmente aceptadas, ej: RENFE.
- Las palabras de jerga y argot si se deben admitir salvo que estén muy generalizadas, ej: HIPPIE
- Se debe utilizar la forma más empleada de llamar a las cosas, lo mismo que la ortografía. Ej: Setiembre en vez de Septiembre, Magdalena en vez de Magdalena.
- Se deben evitar los modos y neologismos en las expresiones.
- Los homónimos (palabra con igual significado) deben distinguirse o bien poniendo entre paréntesis lo que significan o bien con distinto tipo de letra.

Plantas (vegetales o nucleares)

Plantas

- Las denominaciones comerciales no se deben admitir salvo que se conozcan sólo con ese nombre. Ej: Aspirina

Ácido acetil salicílico

- Los nombres propios pueden aparecer como identificadores, lo normal es traducirlos, depende también de si son muy conocidos.

De qué zona se deben extraer los descriptores:

- 1.- Título
- 2.- Subtítulo
- 3.- De los párrafos de entrada del capítulo
- 4.- Palabras destacadas del autor
- 5.- Resumen si tiene
- 6.- Del propio texto

20 / 3 / 2000

PRINCIPIOS QUE RIGEN EL PROCESO DE INDIZACIÓN (Examen)

Los principios se aplican sobre todo en la segunda fase: LA SELECCIÓN

Los distintos autores suelen mezclar unos principios con otros, esto es normal, hay que diferenciarlos muy bien.

Existen cuatro principios fundamentales que llevan parejo otros conceptos íntimamente relacionados, por lo tanto hablamos de PAREJAS DE PRINCIPIOS:

a.- COHERENCIA – CONSISTENCIA

COHERENCIA: Grado de concordancia en la representación de la información esencial en un documento por medio de un conjunto de términos de indización seleccionados por cada uno de los indizadores (de un grupo, un centro, etc...)

La norma UNE 121-91 es muy importante

Existen tipos de **incoherencia**:

1.- Incoherencia interna: es la del indizador. Es aquella que el indizador, indizando un mismo documento en tiempo distinto utiliza descriptores distintos.

2.- Incoherencia externa: se comete entre distintos indizadores indizando un mismo documento, ponen

descriptores comunes pero también descriptores distintos.

Qué mensaje nos manda una **incoherencia elevada**: fallo en el lenguaje. Por lo tanto hay que revisar el lenguaje de la indización

El mundo anglosajón utiliza **CONSISTENCIA**.

Consistencia: es algo que se mantiene en el tiempo. Por ello si un sistema es coherente desde que se inicia, y es coherente en el tiempo diremos que es consistente.

La coherencia es fácil detectarla entre los indizadores o el propio indizador. La coherencia también puede darse en los lenguajes de indización, ha podido comenzar con unas principios y terminar con otros principios, en este sentido es lenguaje también incoherente. Esto se fácil que ocurra con los lenguajes precoordinados y los lenguajes postcoordinados.

Un sistema incoherente no es sinónimo de malo.

Repaso:

Un lenguaje precoordinado: está diseñado para reconocer una cadena de caracteres, no para combinar. Los descriptores se combinan antes de la indización.

Postcoordinado: suelen ser términos que se combinan a la hora de la recuperación. Está hecho para combinar los elementos.

*¿Cómo se logra que la incoherencia no se mantenga en los lenguajes?

Poniendo al día los lenguajes, logramos la máxima coherencia. Exige la presencia de un lenguaje documental, las bases de datos que utiliza un lenguaje libre tendrán más incoherencia que las bases de datos que utilizan un lenguaje controlado.

La coherencia total en un sistema de información es imposible al 100%.

Uno de los factores que hace que sea imposible es el factor intelectual del indizador. Por lo tanto si son varios los indizadores aunque sea mínima la subjetividad que apliquen va a haber incoherencia.

b.- RELEVANCIA Y PERTINENCIA

RELEVANCIA: La capacidad y la adecuación de un término de indización para expresar un concepto determinado mejor que cualquier otro término.

Ese término y no otro es el que tiene una relevancia del 100%, y será el término que se elija como descriptor.

El efecto que se obtiene es recuperar la información de una forma óptima ante una petición de búsqueda determinadas, si se da el descriptor se encuentra bien.

La Relevancia está directamente relacionada con la **PERTINENCIA**

PERTINENCIA, es la misma adecuación término-concepto pero referida a una demanda concreta.

Es decir, yo puedo tener un término relevante y los documentos que recupero son relevante, pero a lo mejor no son pertinente para un señor en concreto.

La norma UNE 121-91, establece 3 factores que determinan la relevancia de un término:

- 1.- El tema
- 2.- La naturaleza del documentos (si es una revista, un artículo, etc...)
- 3.- Las necesidades de información de los usuarios.

El principal criterio de selección de conceptos debe ser su valor potencial como elemento del expresión del tema del documento para su recuperación. Además debe poder identificar todos los conceptos de un documento que tengan valor potencial para los usuarios del sistema de información. Es decir, a la hora de elegir términos debo pensar no sólo en su valor potencial a la hora de la recuperación, sino también pensar en extraer los conceptos que pueden tener utilidad potencial para los usuarios y pertinencia.

c.- LA ESPECIFICIDAD – PRECISIÓN

ESPECIFICIDAD: La Norma UNE habla de:

La Especificidad está relacionada con la exactitud con que un concepto particular está representado por un término de indización (si tengo un concepto muy particular lo indizo también con un término muy particular, si el concepto es general , lo indizo de modo general). Se produce una pérdida de especificidad cuando un concepto particular está expresado por un término que tiene un concepto más general.

Recomienda:

Las nociones deben identificarse o indizarse con el término más específico posible.

Cuanto más específico se indice un documento más relevante será ese documento.

*** Cómo se puede lograr la especificidad en el lenguaje, dado que hay grados de especificidad:**

El grado depende de la forma en la que se representen los conceptos.

Si un término de indización cubre el concepto entero, si un término abarca la totalidad de ese concepto hablamos de un grado elevado de especificidad.

Si es más genérico el concepto, la especificidad irá descendiendo a medida que subimos en el árbol genérico.

A mayor generalidad, menos especificidad.

Ej. Aviones, helicópteros, aeroplanos– se indiza por transporte aéreo

Aquí Lancaster da su punto de vista y dice que es preferible usar varios términos específicos que no uno más general., aunque también se corre el peligro de la redundancia.

Ej: cítricos: naranjas, pomelos, limones.

***¿Qué influencia tiene en el lenguaje documental la especificidad?**

Esta relación es fácil de entender dado que la especificidad la marca el Lenguaje documental.

La especificidad del Lenguaje documental depende de dos factores:

1.- La materia que cubra el lenguaje documental

2.- La precisión

1.- De la materia decimos:

Hay lenguajes documentales más generales o universales, por ejemplo; lo encabezamientos de materia.

Lenguajes documentales más específicos. Cuanto más especializado sea un lenguaje más específico es.

2.- Precisión o extensión del lenguaje.

Cuanto mayor o menor sea el volumen del lenguaje documental empleado, la especificidad será mayor o menor.

Por ejemplo; indizar por palabras clave sin control terminológico, el lenguaje seme dispara, dado que puedo poner tantas palabras como quiera.

* Métodos para alcanzar la especificidad a través de los Lenguajes Documentales.

Se hará estableciendo enlaces sintácticos entre los descriptores.

Se pueden establecer estos enlaces de cuatro formas:

1.- Especificidad de un punto de vista:

Esto es a lo que habitualmente se le llama MODIFICADOR.

Es un término que por si mismo no tiene entidad suficiente para ser descriptor pero que unido a un descriptor complementa, añade o quita información.

Ej: Pronóstico + descriptor

Prevención + descriptor.

Ej:

En un lenguajes precoordinado sería: pronóstico del cáncer de pulmón.

En un lenguaje postcoordinado sería: pronóstico – cáncer – pulmón.

No se deben confundir IDENTIFICADORES con MODIFICADORES.

2.- La especificidad de un vínculo o sistema de nexos:

Esto, el establecer vínculos se suele hacer para evitar falsas combinaciones.

Se establece por una serie de códigos, nexos entre descriptores que vallan juntos.

Ej: política– educación.

Ej: No encontramos un registro cuya cadena de descriptores es la siguiente:

comercio–barco–madera–Japón–puerto–USA.

Entendemos que el documento puede ser:

- a.– Comercio por barco de madera japonesa en puertos americanos
- b.– Comercio por barco de madera americano en puerto japonés.
- C.– Comercio de madera por barco japonés en puerto americano.

Por la cadena que teníamos al principio no puedo saber cuál de esto tres documentos es el que necesito.

Entonces pongo claves

Comercio (A)	Comercio (A)	Comercio (A)
Barco (A)	Barco (A)	Barco (B)
Madera (B)	Madera (B)	Madera (A)
Japón (B)	Japón (C)	Japón (B)
Puerto (C)	Puerto (C)	Puerto (C)
USA (C)	USA (B)	USA (C)
EJEMPLO 1	EJEMPLO 2	EJEMPLO 3

Formas de establecer especificidad a través de nexos.

El indizador dice: Comercio se refiere a barco, Madera va con Japón y Puerto con Usa.

3.– Tercer forma de lograr la especificidad es a través de indicadores de función o indicadores de rol:

Nos dice la función que desempeña un término cualquiera. Es decir, qué función o que rol desempeña un descriptor cualquiera.

Ej: un rol de destino, un rol de producto, etc...

4.– La Yuxtaposición u orden fijo

Es lo más corriente en las bases de datos.

Es el orden por que el que se ponen los descriptores en la cadena documental.

Ejemplo en el primer caso:

Comercio – Barco – Madera – Japón – Puerto – USA

En cuanto yo vea la cadena documental ya se que esa es la secuencia. De manera que al leer la cadena veo que el orden es ese, aunque no veo el sentido.

27 / 3 / 2000

d.– EXAHUSTIVIDAD

La exahustividad está relacionada con la extracción de los conceptos

El indizador no puede por si sólo llegar a la exahustividad ya que no sólo depende de él, sino de factores ajenos, del lenguaje que utilice, el objetivo del centro, de la política, etc...

Lancaster la medida en la que la materia tratada en un documento particular es cubierta en la representación de ese documento, se corresponderá aproximadamente con el número de términos asignados o alguna otra medida del número de puntos proporcionados

La máxima exahustividad es aquella en el que el documento tenga tantos descriptores como temas o conceptos estén dentro del documento.

Una exahustividad del 100% únicamente se logra a texto completo.

Es bastante obvio que la exahustividad en la indización está relacionada con la exahustividad en la recuperación, sin embargo, la exahustividad es inversamente proporcional a la precisión.

Es decir, un nivel alto de exahustividad en indización produce una exahustividad alta y una precisión baja en la recuperación.

Un nivel bajo de exahustividad en la indización produce una exahustividad baja y una precisión alta en la recuperación.

Cómo se consigue mejorar, elevar o eliminar la exahustividad en la indización:

Hay formas para que a la hora de indizar puedan alcanzar el nivel de exahustividad deseado:

1.- Método de ponderación:

El la asignación al término de indización que halla elegido de un valor positivo asociado a ese termino (con lo cual yo pondero el valor que deseo)

Ejemplo:

Si me encuentro 10 descriptores que me hacen alusión a 10 temas ¿cómo se yo que descriptores son los que más hacen alusión al tema deseado?

Lo hago por medio de esa ponderación

A la hora de indizar doy una clave para saber a la hora de recuperar qué descriptores están tratados con mayor o menor grado en la indización.

La escala que se utiliza son del 1 al 7 son lo que se suele denominar: Descriptores principales y descriptores secundarios. Es una manera de ponderación.

Los descriptores principales hacen referencia a los temas que más se han tratado.

Los descriptores secundarios son los que tocan los temas pero de manera más tangencial.

A la hora de recuperar la información puede solicitarse al sistema que me recupere sólo los descriptores principales o sólo los descriptores secundario, o ambos a la vez.

Hay tres problemas que se conocen respecto a esta indización ponderada:

1.- Elegir correctamente la escala de valores

Si nos movemos en escalas del 1 al 2 es fácil, pero en escalas del 1 al 7 no es tan fácil, además influye la subjetividad del indizador.

La ponderación ayuda a la mayor pertinencia en la indización.

El elegir correctamente la escala de valores está relacionado con la asignación de los valores.

2.- Aplicación del propio indizador.

3.- El tiempo, lógicamente indizar de una forma ponderada incrementa bastante el tiempo de indizar, con lo cual influye en el coste.

La ponderación está directamente relacionada con la ESPECIFICIDAD.

Alvarez Osorio Acto que habla claramente sobre la ciencia de la Documentación.

La especificidad y la exhaustividad son dos magnitudes inversamente proporcionales.

A mayor exhaustividad menor especificación. Son principios que tiene que marcar el indizador.

Se identifica la exhaustividad con el número de descriptores que acompañan a un documento.

En la NORMA UNE 521 dice que es absurdo que el número de descriptores nos hablen de la exhaustividad. Más bien hablaremos del tipo de lenguaje.

Los descriptores en el lenguaje precoordinado: al se un lenguajes, se van a utilizar menos términos para indizar el contenido ya que los precoordinados son más largos.

Mientras que en una postcoordinación, se exige mayor número de términos.

Hay por ejemplo bases de datos que con 1 o 2 descriptores nos han indizado el documento. Por lo tanto, **es falso que la exhaustividad dependa del número de descriptores.**

2.- **El último principio en indización es el Modelo de Entropía**, lo que se denomina **ahorro documental**, hay que dar la mayor carga posible con los descriptores utilizados, el menor número de términos posibles.

Ese descriptor es capaz de dar mucha información con menos descriptores.

Cuanta más carga semántica tenga ese descriptor mejor.

TEMA 7.-

PRODUCTOS QUE GENERA LA INDIZACIÓN

En este sentido, los ordenadores han dado mayores prestaciones a lo que ha sido la indización no sólo respecto al producto. Los productos que genera el ordenador son una serie de índices y autoíndices.

Los índices más conocidos son los PERMUTADOS, lo realiza el ordenador automáticamente a partir sobre todo de las palabras de los títulos indizados.

Los índices más utilizados son:

KWICK: Key words in context, presentan las palabras dentro del contexto

KWOCK: Key words of context, las palabras fuera de contexto.

Los KWICK presentan las palabras clave alfabéticamente con permutación, es decir, las palabras significativas al título se reencuentran todas en el mismo lugar unas seguidas de otras.

Aparecen todas las palabras significativas del título, lo normal es que se alfabeticen. Un término aparece tantas veces como veces aparezca en el título del documento.

Ejemplo: Títulos

- 1.- Enseñanza de los LENGUAJES DOCUMENTALES en documentación
- 2.- Organización de los LENGUAJES DOCUMENTALES
- 3.- Coordinación de los LENGUAJES DOCUMENTALES en un texto
- 4.- Uno de los LENGUAJES INFORMATICOS en una base de datos.
- 5.- Establecimiento de los LENGUAJES UNIVERSALES en Biblioteconomía.

Cada título tiene un número de referencia. EJ . el 1, 2, 3, 4, y el 5.

Los KWOCK no presentan las palabras significativas, sino que esas palabras aparecen fuera del título, te alfabeticen a partir de las palabras que hacen de descriptores, así tenemos:

LENGUAJES DOCUMENTALES

Enseñanza de los

Organización de los

Coordinación de los.....

LENGUAJES INFORMÁTICOS

Uso de los

LENGUAJES UNIVERSALES

Establecimiento de los

Para indizar se suele utilizar más el KWICK porque el KWOCK supone que ya conoces los términos.

El KWICK se utiliza más para indización,

el KWOCK para recuperación.

Otro tipo de índices son los ÍNDICES ACUMULATIVOS, acumulan la información en un periodo

concreto.

Si yo tengo 4 índice de todo un año, este acumula todos los índice y los saca en uno

ÍNDICES DE CITAS, a pesar de que son una herramienta muy específica de bibliometría es necesario saberlo. Son índices de autores pero en estos índices no sólo se cita el autor y la lista de documentos que este autor ha publicado, sino que a la vez en ese índice aparece la lista de los documentos en los que ese autor ha sido citado.

Estos índices tienen un índice de materias permutado y también viene un índice con las fuentes donde aparecen los trabajos citantes ordenados por autores que han citado y las instituciones de donde procede esas citas (dónde se ha publicado, etc...)

Con esto puedo saber que relevancia tiene ese autor.

Los dos índices de citas por excelencia son:

* El índice más conocido es el SCI Science Citation Index, que lo publica el ISI Information Social Institute.

ÍNDICES BIBLIOGRÁFICOS, contienen las referencias bibliográficas de un conjunto de documentos, se ordenan por cualquier característica (tanto de forma como de contexto) de ese documento.

Son muy rápidos de aparecer y en este sentido fácil de llegar al público, pero no es fácil saber la información de ese título. Presentan información de servicio de alerta.

Hay veces que incluyen las palabras claves, pero no es lo habitual

ÍNDICE DE CONTENIDO, Reproducen la parte de las revistas seleccionadas de la página del índice, también acompañan un índice de materias y otro de autores son fáciles de preparar.

Su utilidad en el servicio de alerta. Con una ojeada, la gente a la que le interese un tema sabe donde lo puede encontrar.

* NO SON ÍNDICES, pero no dejan de ser el producto por excelencia de la indización, son las **BASES DE DATOS**.

Definición de bases de datos dada por la FID: Conjunto de datos homogéneos ordenados de una forma determinada que se presentan normalmente legibles por ordenador.

3 / 4 / 2000

TEMA 8.-

LOS TESAUROS

Es la herramienta por excelencia en cuanto al logro de la mejor indización.

TESAURO:

Herramienta por excelencia a la hora de la indización que comprende un conjunto controlado de términos entre los cuales se establecen relaciones jerárquicas o de asociación con las palabras del lenguaje natural, lo cual se aplica a un campo particular del conocimiento.

1.- Los términos que comprende el tesoro tienen una significación única que implican el control de la polisemia, sinonimia, etc.... Carecen de cualquier tipo de ambigüedad.

2.- Con la herramienta de un Tesoro se logra la exhaustividad y la pertinencia.

Para definir los tesauros el Plan General de Información (PGI) tiene en cuenta 2 puntos de vista:

a.- La Estructura

b.- La función.

A.- Según LA ESTRUCTURA:

Un Tesoro es un vocabulario controlado, cuyos términos son muy dinámicos y evidentemente, guardan entre sí relaciones de tipo semántico y jerárquico. Y desde el punto de vista de la estructura se aplican estos términos exhaustivamente a un área determinada de conocimiento.

B.- Según SU FUNCIÓN:

Desde el punto de vista de la función el Tesoro es un instrumento de control terminológico que sirve para convertir el Lenguaje Natural en Lenguaje Controlado.

Los Tesauros son una obra inacabada, es una cosa viva, tiene que adaptar los lenguajes naturales a lenguajes controlados, su tamaño estará en función del tiempo que tenga ese tesoro.

Se considera que un tamaño es idóneo o recomendable cuando en general contempla unos tres mil términos.

Lo normal es que los Tesauros sean especializados, que se dediquen a un área determinada.

El primero de los Tesauros que aparece es en el año 59 (era de ciencias).

El Tesoro aunque no es un producto de la indización, si es una herramienta útil para la indización.

La unidad básica del Tesoro es el DESCRIPTOR, pero el hecho es que no es tan importante elegir ese descriptor, sino la elección de la relación de esos términos.

A.- Según LA ESTRUCTURA:

La estructura de un Tesoro la comprenden cinco partes diferentes:

1.- Descriptores, identificadores y modificadores.

2.- En la presentación formal de los descriptores

3.- La presentación formal de las palabras, aquí lo que hay que tener en cuenta es su ortografía, si esta en forma sustantiva, su número, si está en singular o plural, abreviaturas, siglas, etc...

4.- Presentación de las ambigüedades, no se tratan estas ambigüedades entre los términos. Las ambigüedades se controlan a través de la polisemia, sinonimia, antonimia, homonimia, etc...

El tesoro tiene que explicar como ha hecho para evitar esa homonización, polisemia, etc...

Tiene que explicar de cada término cuales son sus diferentes significados a través de:

- a través de notas de aplicación o función, se da la definición de los términos.
- o también a través de las traducciones de los términos.

5.– Establecimiento de las relaciones entre los descriptores. Estas relaciones pueden ser de cuatro tipos:

a.– Relaciones de equivalencia o preferencia

Son los sinónimos, las palabras con distinta escritura (v,b, etc...) y las traducciones, siglas y abreviaturas. Se remiten por el USE O USADO POR.

b.– Relaciones alternativas

VÉASE O VÉASE ADEMÁS (o viene de)

c.– Relaciones de jerarquía

Se mueven en dos aspectos:

C.1.– Relaciones genéricas

C.2.– Relaciones específicas o partitivas

C.1.– Relaciones genéricas: Se refieren al género o a la especie:

Constan del término amplio, del término genérico y del término específico.

La información que producen va disminuyendo cuanto mayor es el término, más información.

C.2.– Relaciones específicas o partitivas: Se refieren a las partes y al todo. Hacen referencia sólo al término específico.

d.– Relaciones asociativas:

Tienen carácter horizontal o de vecindad, en estas relaciones lo que se establece son relaciones entre dos términos, la antonimia (coordinación entre dos términos), la coordinación genética (padre-hijo).

Causa-efecto.

Instrumental: mando a distancia con respecto a la televisión.

Materiales: Jersey con lana

Similitud:

Todo ello se expresa a través de los términos relacionados TR

Estas relaciones se generan en los Tesauros.

Para elaborar un Tesauro se tiene que tener en cuenta:

- No duplicar tareas.
- Saber si ya existe alguno sobre esa materia
- Conocer la normativa internacional, la normalización de tratamiento del Tesauro. UNISIT, ISO, etc...
- Se debe estudiar antes de realizar un Tesauro el sistema de recuperación que se va a utilizar. El lenguaje de indización si va a ser postcoordinado o precoordinado, etc...Lo normal es utilizar el postcoordinado.
- Identificar cuales son las áreas de las que trata el Tesauro. Delimitar el campo de actuación del Tesauro.
- Saber de que área científica va a tratar.
- El tiempo que nos va a llevar hacer todo esto ya que está en relación con el coste.
- Sobre el coste, saber con qué infraestructura informática y profesional se cuenta.
- Saber a qué usuarios va dirigido. El tipo de usuarios va a configurar el tipo de vocabulario, si hay profundidad o no, tipo de lengua a emplear, etc...

El Tesauro debe pasar por una serie de fases antes de llegar a la venta.

1.- Recogida de términos:

Se recogen de varias formas, lo normal es a través de la indización de los documentos.

Se calcula que esa recogida de términos debe llevar al menos un año de indización.

Una vez recopilados estos términos se pasa a la segunda fase.

2.- Fase de depuración:

3.- Establecimiento de los descriptores:

Es hacerlo a través de los términos de indización o también ayudarnos de herramientas (diccionarios, enciclopedias, bases de datos, etc...)

Una vez que ya se tienen los descriptores hay que establecer las jerarquías.

4.- Jerarquizar esos descriptores:

Ponerlos de mayor a menor.

Una vez hecho todo esto,

5.- Agrupar los descriptores elegidos por grandes grupos o familias.

Después se pueden establecer las relaciones entre ellos, lo normal es que partan de las propias familias.

6.- Presentación del tesauro

Es la primera puesta en marcha del tesauro. Una vez editado de forma impresa.

7.- Evaluación del tesauro:

Se recomienda que se evalúe el tesauro con un periodo mínimo de 6 meses o bien se recomienda que se índice como mínimo unos 1000 documentos.

Los indizadores deben recuperar la información eligiendo los descriptores de ese Tesauro, y en base a lo que recuperen deben hacer pruebas para ver si es relevante la información que recibe de ese tesauro.

El indizador debe medir tanto la precisión como la exhaustividad, en base a ello habrá que hacer las modificaciones necesarias.

8.- Corrección del Tesauro

Habrà que añadir o quitar temas,

9.- Editarle, sacarle y mantenerlo.

Lo normal en el mantenimiento de los Tesauros es que cada cierto tiempo se saquen anexos al tesauro que ya está editado para ver los nuevos términos o los suprimidos.

Cuando lo anexos son mucho, se edita un nuevo Tesauro.

TIPOLOGÍA DE TESAUROS

Se puede establecer la tipología de 4 maneras:

1.- Según sus características formales:

1.a.- Tesauro que sólo tiene una parte alfabética,

1.b.- Y Tesauros que presentan una parte jerárquica.

1.c.- Lo normal son los Tesauros mixtos, es decir, presentan las dos partes, tanto la alfabética como la jerárquica.

2.- Según su contenido:

2.a.- Macrotesauros:

Abarcan mucho, son pluridisciplinarios, y sus características son de mucha generalidad ya que abarcan muchas disciplinas.

2.b.- Microtesauros:

Son partes abreviadas de los macrotesauros, o de uno ya existente que se usa poco.

2.c.- Sectoriales o especializados:

Sólo contemplan una o dos disciplinas, de tal manera que su grado de especificidad es muy grande.

2.d.- Mono o multilingüe,

Dependiendo si tienen una o varias lenguas.

3.- Según su objetivo

Según lo que se persiga.

4.- Según su cobertura geográfica:

Internacional, institucional, de centro, de red, regionales, etc....

CARACTERÍSTICAS GENERALES DE UN TESAURO

- Debe cubrir entre 5 a 50 dominios o familias.
- Debe intentarse que sea compatible con otros.
- También es deseable que una de sus características a controlar sea el tamaño, que no sea muy excesivo para poderlo manejar.
- Lengua a contemplar, la que se va a usar.
- Tipo de relaciones: – que sólo tenga un tipo

– cuantas más relaciones mejor.

6. Así como sus relaciones formales en cuanto a relación, listas alfabéticas,

Gráficas. (Dibujos, etc...)

PARTES DE UN TESAURO

- Tiene o debe tener una portada, donde venga el título, autor y fecha .
- Sumario: debe incluir las listas jerárquicas establecidas.
- Introducción: aquí debe explicar cual son las pretensiones que se buscan, metodología, etc... Se debe incluir la cobertura, la metodología, criterios utilizados en el tesauro, fuente utilizada para la elección de los términos, etc.... Objetivos que se han planteado, innovaciones si es que las hay, modo de uso, es decir todo cuanto se pueda.
- Lista estructurada alfabética, es lo que se denomina el Tesauro alfabetizado. Se debe expresar aquí las relaciones horizontales, que se exponen o letra a letra o palabra a palabra si hay sólo términos compuesto o simples. Se alfabetizo por los dos términos o sólo por el primero de ellos.
- Establezco una lista con las jerarquias, que es el tesauro clasificado. Se especifica la relación de jerarquía ¿Cómo? Se establece en el Tesauro estas jerarquías o bien por sangrados o a través de gráficos. Hay tesauros que tienen las dos (en forma de árbol y en forma de flechas).
- Índices auxiliares: me ayudan a saber cómo están establecidos estos términos. Tipos: Kwick, Kwock, índice por grupos de descriptores y pueden ser de nuevos descriptores. También puedo tener un índice que me diga cuál son los nuevos descriptores incluidos, suprimidos, modificados, etc... Listados geográficos, etc....

FUENTES TERMINOLÓGICAS DE LAS QUE SE DEBE NUTRIR UN TESAURO:

- Por un lado de una forma manual, de los descriptores indizados.
- Tratamiento de texto.
- Los sistemas automatizados de gestión de un Tesauro.
- De manera mixta, combinando diferentes tipos de fuentes terminológicas.

RELACIONES ENTRE DESCRIPTORES EN UN TESAURO:

1.- EQUIVALENCIA

USE !

UF = (used for) (usado por).

2.- ALTERNATIVAS

S (see) :

SF (see from), * , (viene de)

3.- JERÁRQUICAS

3.1.- RELACIONES JERÁRQUICAS GENÉRICAS

TA (termino amplio), "

TG (terme generique)

BT (broader term)

3.2.- RELACIONES JERÁRQUICAS ESPECÍFICAS

TE (term. Especif), <

ST (specifique terme)

NT (narrow term)

4.- ASOCIATIVAS

TR (termino relacionado), __

TA (terme asocié)

RT (related term)

10 / 4 / 2000

TEMA 9

LA CALIDAD DE LA INDIZACIÓN

9.1.- La importancia que tiene la calidad de la indización:

La calidad no es un fin en si misma, hay que evaluar esos resultados, es cuestión de recuperación.

Hay dos cosas a tener en cuenta sobre la importancia que tiene la calidad de la indización:

a.- Los productores de bases de datos han hecho poco caso en este sentido, si es mejor la base de datos, mejor será la calidad de indización en cuestión de recuperación.

b.- El usuario demanda mayor calidad en este tipo de productos electrónicos que la que han demandado en cuestión de productos impresos.

Uno de los autores que más ha escrito sobre el tema de calidad es SOERGEL (es un autor contemporáneo). No sólo ha escrito, sino también ha aplicado. Dice es necesario que los productores de información o de bases de datos se han estas tres preguntas

- 1.- Si la indización mejora el rendimiento
- 2.- se estas mejoras aumentan el precio de la base de datos
- 3.- qué características de la indización son más importantes de efectuar o repetir.

El dice que al haber pocas exigencias no ha habido tiempo de efectuarlas, aunque cada vez son más las que se están llevando a cabo.

9.2.- Aspectos que influyen sobre la calidad de la indización:

1.- Factor relativo al propio indizador:

Si el indizador tiene conocimientos del contenido temático que maneja, indizará mejor. También aquí influye que el indizador conozca los intereses de los usuarios. Cuanto mejor conozca esas necesidades más fácil será plasmarlas.

La experiencia del indizador.

Al indizador se le suponen una serie de características que hacen que sea mejor, como facilidad de lectura, capacidad de extracción de descriptores rápidamente, etc....

2.- Relacionadas con el vocabulario de la base de datos:

Cuanto más específico, mayores matices se le da a la indización, es decir, siempre es deseable que los descriptores sean específicos.

También mejora la calidad el establecimiento de enlaces sintácticos entre ellos (los descriptores). Este es uno de los factores que mejora la indización y aumenta el coste.

Si los términos son ambiguos o imprecisos harán que esa indización sea peor, por lo que dificulta la elección de los términos.

El tamaño y calidad del vocabulario a emplear. Cuanto mayor sea, mayor exhaustividad, pero puede haber imprecisión.

Disponer de instrumentos auxiliares, diccionarios, enciclopedias, etc...

3.- Factores relacionados con el propio documento:

El propio contenido temático de un documento también hará que la indización sea más o menos difícil.

El idioma del documento.

La calidad de exposición del documento,

La extensión del documento.

Los elementos del documento, si tiene resumen, epígrafes, etc.. es más fácil la indización.

d.- Factores relacionados con el propio proceso de indización:

Por ejemplo, el proceso de extracción es más fácil de indizar, si extrae palabras del propio documento. Mientras que hay otros tipos de indización que dificultan la calidad.

El uso de un lenguaje postcoordinado o precoordinado.

El precoordinado es un lenguaje más general .

El postcoordinado mejora la indización, la recuperación es mejor.

Si el centro dispone de unas reglas en cuanto al propio proceso de indización también esto facilitará la calidad en la indización.

La exhaustividad del sistema influye en la calidad. Cuanta más exhaustividad, más coste.

5.- Factores ambientales:

Las condiciones en las que un indizador realiza su tarea afecta al propio trabajo, por lo tanto influye en la calidad.

9.3.- Medidas que se aplican en ciertas características de la indización a la hora de ver la calidad en la indización :

a.- El grado de coherencia

b.- El principio de relevancia

c.- Grado de especificidad

d.- Exhaustividad.

a.- El grado de coherencia

Al medir la calidad en la coherencia también habría que hablar de consistencia.

La manera de medir la coherencia en indización es medirla en base a los indizadores.

Incoherencia externa: la que existe entre distintos indizadores indizando un mismo documento.

Incoherencia interna: es la del propio indizador.

Dividir el número de descriptores comunes A- B

Para medir la incoherencia externa : _____

Partido por el número de descriptores diferentes.

El resultado nos dará si esta incoherencia se puede dejar o hay que modificarla porque es elevada.

También se puede.

Dividir el número de documentos a los que el descriptor fue asignado por los indizadores.

Para medir la incoherencia interna: _____

Partido por el número de documentos a los que el descriptor fue asignado por uno de ellos

b.- El principio de relevancia

Exactitud en que un concepto que aparece en un documento está representado por un término.

El grado de relevancia se debe considerar desde distintos puntos de vista, dado que la relevancia es muy difícil encuadrarla. Para un autor un término es muy importante, y para otro autor no es tan importante ese término.

JO y AJIFUQUF , ellos dicen que la relevancia se puede medir a través de valores de discriminación.

Para calcularla:

número de registros que recupera asociados a un descriptor

Número total de documentos en la base de datos

Se calcula que el valor de discriminación correcto es aquel que se aproxima al 0,05, puesto que otro corresponde al 5% de documentos de cualquier base de datos.

Estos valores se empezaron a calcular sin tener en cuenta el tamaño de la base de datos, pero esto no es correcto, habría que hacerlo teniendo en cuenta el tamaño de esa base de datos.

Con los valores de discriminación se analizan:

- el grado de especificidad, y generalidad de un documento.
- la especificidad está relacionada con la relevancia.
- si los términos explican bien el contenido de los documentos.
- se ve también si tengo términos que no son relevantes, es decir, si han sido recuperados por otros , con otros sinónimos más relevantes.

c.- Grado de especificidad o principio de especificidad.

Que indiza con términos genéricos o específicos.

La especificidad en un sistema la marca el propio lenguaje de indización. El propio lenguaje nos marca a qué nivel de especificidad me puedo mover.

Cómo se mide la especificidad de una indización.

En los lenguajes que yo conozco y utilizo encuentro que con un lenguaje postcoordinado encuentro un mayor nivel de especificidad

La especificidad también nos mide la propia capacidad de coordinación de los términos.

d.- Exahustividad.

(Número de conceptos que incluye un documento, de qué manera está representado por un número de documentos).

Ej: Un bloque de 10 temas tendría 10 descriptores.

La fórmula de hallar la exahustividad es poco fiable, se halla sacando en una base de datos el porcentaje de descriptores que se crean en los documentos.

Ej:

100 documentos

En uno han utilizado 10 descriptores, en otro 7, en otro 3, y así sucesivamente hasta llegar a los 100. Si hallamos la media nos dice el nivel de exahustividad, en este caso puede ser de 6.

Lo que si te indica es si existe una política en el centro (en un centro se mueven entre valores de 4 y 6 descriptores) (en otro centro hay documentos que se indizan por 1 o por 10 descriptores), esto nos indica la política.

Según el tipo de lengua.

Otro de los factores que influyen es la extensión del documento, cuanto más extensión más probable que se utilicen más descriptores.

Del propio sistema de indización, según la exactitud del lenguaje, de la indización que tenga el sistema, etc... De todo ello depende la calidad.

8 / 5 / 2000

TEMA 10

LA AUTOMATIZACIÓN DE LA INDIZACIÓN

Es un tema envuelto entre polémica entre sus defensores y retractores, ya que no se ponen de acuerdo los investigadores.

Con respecto a los defensores, alegan que es una ayuda para el indizador, es más objetiva ya que no corren el peligro de la subjetividad. Es más fiable, ahorra tiempo.

Entre los detractores no se ha logrado un éxito en los resultados, el coste es grande, elimina la subjetividad pero no capta algunos matices. No se justifica el coste con los resultados.

No hay una trayectoria muy larga dado que empieza en los años 50. Se están haciendo logros, también es la mejor de las prestaciones con respecto a la ordenación, está mejorada supone avances con respecto a la indización automatizada.

La Indización automatizada es un proceso complejo ya que en su desarrollo intervienen varias disciplinas. La aparición de la lingüística computacional, supone aplicaciones de las ciencias de la computación a las estructuras del lenguaje.

El introductor de la lingüística computacional es CHOUSKY que supone un paso más en cuanto a la Indización automatizada.

La INDIZACIÓN AUTOMATIZADA: es la aplicación o el uso de ordenadores para extraer y aplicar términos de indización sin que intervenga el ser humano. Esta definición se debe a STEVENS.

Desde los comienzos de investigación hubo dos procedimientos en cuanto a la indización automatizada:

- por un lado procedimientos lingüísticos
- y procedimientos matemáticos.

Actualmente se tiende a conjugarlos.

Los procedimientos matemáticos también se denominan métodos estadísticos, y a los procedimientos lingüísticos también se les denomina no matemáticos.

Estados Unidos y también Francia han sido los que más han investigado, en España hay un retraso en la investigación.

Respecto a estos dos métodos, a su vez se dividen en varios sistemas de indización automatizada.

1.– Dentro de los métodos matemáticos, no lingüísticos o estadísticos:

Son los primeros en aparecer, son los más fáciles de utilizar.

Encontramos:

1.a.– **Método de frecuencias**, se debe a LHUN.

Se basa este método en que la mayor o menor frecuencia de aparición de los términos de un documento está relacionada con su utilidad a la hora de indizar.

Que un término aparezca muchas o pocas veces tiene que ver con su importancia.

* Términos genéricos: aparecen muchas veces, pero producen menos precisión.

* Términos de frecuencia baja: termino específico que implica una baja exhaustividad.

Este método de frecuencias dice que para elegir un término de indización no son buenos ni los muy genéricos ni los poco genéricos. Los más idóneos para indizar son los que tengan frecuencias medias.

Las frecuencias de aparición de un término son muy significativas en cuanto a cómo representan el contenido.
No es un método muy fiable.

1.b.– Aparece el segundo método casi simultáneamente con el anterior, su variación sobre el primer método se debe a DAMEREAU.

El cambio que introduce es que él dice que él habla de frecuencias absolutas de un término de palabras. El dice que no sólo hay que fijarse en esta frecuencia absolutas, sin que lo que realmente es importante es la frecuencia relativa (es el número de veces que aparecen, dividido entre el número de términos que aparecen en la base de datos).

Dice que esto se ajusta mucho a la realidad. Para indizar de una forma automatizada hay que hacer este paso, establecer cuál es el umbral necesario y escoger como término de indización aquellos que se muevan en un número de frecuencia relativa media.

1.- c.- ROSENBERG, introduce el segundo método, es el MÉTODO DE PROBABILIDADES.

En lo que consiste es que Rosenberg confecciona una lista de términos de indización basada en las probabilidades en base a sus frecuencias de aparecer en un documento.

Calcula los términos asociados a est término elegido, es decir, empieza a actuar como un tesoro, no indiza sólo este método sino que introduce la idea de tener los términos asociados.

Es una ayuda real al indizador, ya que cuando está indizando se sirve de esta lista para saber cuales son los términos asociados a los que él ha indizado.

1.d.- MÉTODO DE BOOKSTEING O (Clustering)

Cluster es grupo, agrupación.

En esté método el procedimiento es el siguiente: Está basado en la agrupación de palabras en base a la capacidad para transmitir el tema de un documento.

Lo que hace es que está basado en la frecuencia de aparición de palabras.

El método consiste en estudiar (frecuencias de aparición de las palabras, y no sólo estas, sino también las frases). Cuando estudia la frecuencia de aparición de las palabras se fija también en la posición en la que se encontraban dentro de las frases.

Según su posición, se supone que los términos sean mayor o menor capaces para expresar el tema, el contenido.

Es decir, empieza a introducir la idea de agrupar no sólo palabras sino también frases. Se fija también en el contenido lingüístico.

1.e.- MÉTODO DE SALTON (utiliza el MÉTODO DE DISCRIMINACIÓN)

G. SALTON entiende por valor de discriminación la capacidad que tiene un término para identificar un número determinado de documentos.

Es decir, si yo tengo un término capaz de recuperar X documentos, su valor de discriminación es alto. Por lo tanto, calculo el valor de discriminación y a los términos les aplico un peso según su valor, entonces, si el valor de * es alto, quiere decir que el término es muy general, si es bajo el término es específico.

Los que se prefieren serán los que tengan el valor de discriminación medio, es de los más idóneos.

El valor de discriminación bajo indica escasa exhaustividad.

1.f.- Método que lo introduce SALTON, es el MÉTODO DE RELEVANCIA

Está basado en la frecuencia de aparición de los términos.

Hace distinción entre los que aparecen en un documento relevante y los que aparecen en un documento no

relevante. Hace una distinción cualitativa.

1.g.– MÉTODO DE IMITACIÓN A LA INDIZACIÓN HUMANA

Este método trata de aprender cual son los términos de indización que atribuiría un indizador humano.

Para ello el programa maneja un grupo de términos que previamente han sido elegidos por un grupo de indizadores humanos y calcula normas de asociación entre estos y los nuevos términos a indizar.

2.– Métodos lingüísticos

La segunda forma de indización automatizada la denominamos métodos lingüísticos, están basados en el procesamiento del lenguaje natural.

Surge en los años 50 y está ligada a la lingüística computacional.

Los trabajos teóricos los encabeza CHOUSKY. Están encaminados a la formalización del lenguaje, mientras que los trabajos prácticos se encaminan a la traducción automatizada que es la que mayor éxito tiene al principio.

A finales de los 60 se pasa de la mera comprensión del lenguaje a que los estudios se dirijan al tratamiento del lenguaje, a los procesos sintácticos (estructuras sintácticas).

Es ahora cuando comienzan los trabajos de indización automatizada, es decir, los trabajos sobre este tema. Los primeros trabajos se dirigían al estudio morfológico solamente. La Indización automatizada exige métodos lingüísticos que puedan combinarse con los métodos de frecuencia, van a ser capaces de extraer los contenidos de los documentos. O sea, pasamos de la morfología a la sintaxis. Estamos pues en la organización sistemática. Si esta indización es difícil, la sistemática no lo es menos, ya que no saber diferenciar entre sinónimos, etc....

(Se exige un análisis sintáctico y con ello un método lingüístico, los cuales pueden combinarse con los métodos de frecuencias para solucionar los métodos matemáticos que la falta de precisión, etc... para ajustarse más a la extracción del contenido.

Se pasa de la morfología a la sintaxis).

– Este sistema no sólo tiene en cuenta los conceptos y sus asociaciones, indización semántica– sino que además se fija en la morfología.

VENTAJAS E INCONVENIENTES:

<u>VENTAJAS de los métodos matemáticos</u>	<u>VENTAJAS de los Métodos matemáticos.</u>
1.– Facilidad de aplicación	1.– Son capaces de eliminar ambigüedades del lenguaje.
2.– Se consigue una exhaustividad alta.	2.– Retienen sólo las palabras significativas dentro del contexto.
3.– Sirve de ayuda al indizador.	3.– No son capaces de normalizar el contenido del texto.
4.– Tiene una relativa coherencia.	
<u>INCONVENIENTES de los métodos matemáticos</u>	<u>INCONVENIENTES de los métodos lingüísticos:</u>

1.- No reconocen relaciones semánticas (no son capaces de saber que es una sinonimia)	1.- La metodología es bastante compleja.
2.- Ni reconocen términos compuestos, entonces lo que hacen es que contabilizan cada término por separado, entonces la frecuencia contabilizada está mal.	2.- Hay menor disponibilidad a el mercado de programas o aplicaciones.
3.- Tienen un algo grado de ambigüedad.	3.- Requieren gran capacidad para el almacenamiento.
4.- Precisión baja.	4.- Los resultados tampoco son mejores que en los métodos matemáticos o estadísticos.
5.- No pueden normalizar términos.	

15 / 5 / 2000

TEMA 11

RESUMEN DOCUMENTAL

Es una técnica documental cuyo auge es reciente, su verdadera importancia es reciente.

El resumen documental aparece cuando los escribas empiezan a hacer unos resúmenes de las tablillas fuera de estas tablillas. Con el Impero Romano la Iglesia se encarga de hacer estos resúmenes, en los márgenes de los textos. El resumen documental va parejo al incremento de la información. En el Renacimiento se hacen más resúmenes, los movimientos intelectuales son mayores y por eso es mayor el auge.

La mayor dimensión se alcanza en el siglo XIX, la aparición de revistas científicas trae consigo esos resúmenes. En la segunda mitad del siglo XX sobre todo es cuando aparecen revistas de resúmenes, la primera es en los años 60 cuando empiezan las ciencias y tecnologías, luego van apareciendo distintas materias en esta disciplina hasta que aparecen más tarde en otras ciencias.

Los resúmenes, su principal misión es ser un anticipo del contenido de los documentos originales, es importante para el control, transferencia de información, tanto para autores como para usuarios.

Concepto de resumen:

El concepto de resumen según varios autores como por ejemplo Ottel es el análisis bajo forma suscrita de lo que se contiene en los documentos.

Chaurmier: Es la operación por la que se arregla el contenido de un documento y se representa por un cierto número de operaciones que expresan la sustancia.

Mijailov dice que el resumen es una breve exposición del contenido de un documento en el que se indica el carácter del mismo, los métodos de investigación y los resultados.

López Yepes: es una técnica antigua usada en el trabajo científico para reducir el contenido general enunciado.

José Antonio Moreino: Es una actuación sobre el contenido de los documentos para minimizar la abundancia de información contenida en ellos. Tras ello, el mensaje del contenido original se transforma en otro, es decir, origina otro documento del documento primario.

María Pinto Molina: habla más del proceso de resumir, puesto que abarca operaciones en cadena que indica la metamorfosis que experimentan los documentos textuales desde su estado original, desde su estructura léxico-lingüística hasta su estado macroestructural el cual se refiere a la estructura lógico-semántica.

Es decir, desde su estructura léxico-lingüística hasta lógico-semántica. Dice que el secundario es más completo.

La función:

La función del resumen es informar al usuario de una información ya existente, es como un filtro que controla toda la información que ya posee el centro donde está el documento primario.

El resumen documental es algo ya muy habitual, cada vez es más habitual indizar un resumen, es decir, extraer del resumen los descriptores, por eso su importancia.

FACTORES EN EL PROCESO DE RESUMIR:

Son de dos tipos:

1.- **Textuales:** son factores lingüísticos, afectan a la forma, palabras, estructuras , frases y al significado.

Los factores extralingüísticos afectan al contexto, lógica y psicología.

2.- **Documentales:** afectan al resumen desde el punto de vista documental. Existen 3 tipos:

a.- Las condiciones de producción del resumen, afecta a los recursos del centro.

b.- La propia forma del resumidor.

c.- Exigencias del propio usuario

OBJETIVOS DE LA OPERACIÓN DEL RESUMEN:

- Dar a conocer las aportaciones científicas de los documentos científicos. Es resumen debe contener todos los términos que se van a utilizar en la recuperación.
- Proporciona al usuario la valoración del documento original, a partir de su lectura el usuario decide si lo lee o no.
- Facilitar la comprensión del documento, porque, en poco espacio hay que explicar el contenido del documento.
- Servir de sustituto al documento original.
- Se utiliza en procesos automatizados de indización.

METODOLOGÍA

Distintos autores hablan de distintas metodologías.

Pasos a seguir:

1ª fase: PREANÁLISIS: 3 pasos:

a) Selección de documentos:

La llevan a cabo los centros de documentación e información. Los indicadores que miden estos documentos son los siguientes:

- Razones de economía
- La naturaleza del propio documento
- si es accesible o no
- su aceptabilidad
- su soporte
- la calidad tanto de la publicación como la del autor
- los intereses de los usuarios
- el material

b.- Definir los objetivos:

Campos a cubrir, condiciones de trabajo, tipos de usuarios.

c.- Delimitación del proceso

Es exigible el análisis del texto y la sintaxis del mismo y por supuesto la evaluación de los resultados.

2ª fase: VARIABLES ANALÍTICAS: 2 pasos

a.- El objetivo del resumen

Cuál es el objeto del resumen que suele ser el texto, el documento científico en sentido amplio. En cuanto otro tipo de documento no son objeto de las técnicas documentales que hemos visto.

b.- Determinar cuáles son las partes importantes del texto

Las palabras con mayor frecuencia se han de detectar para hacer el resumen, se considera que el texto es la parte pasiva, la parte activa es el resumidor en sí, el resumidor es activo porque toma decisiones, este debe tener una serie de actitudes.

Se discute quién debe ser el resumidor, esto depende de la persona que lo diga.

Se le exige que tenga conocimientos de la materia. El autor del documento puede que sea y es el que mejor maneje la materia pero a lo mejor no sabe como resumirlos, porque no sabe técnicas documentales. Y también puede pasar al revés.

Al profesional se le exigen ciertas características como: capacidad de redactar bien, que sea hábil, eficaz, que domine la tecnología de resumir, y debe disponer de un tiempo razonable para resumir.

3ª fase: LAS ESTRATEGIAS

La más importante es la lectura del texto, es un acto de perfección textual que desemboca en la comprensión del mensaje (actividad mental).

4ª fase: ANÁLISIS DEL CONTENIDO

Aquí, la protagonista es la lingüística porque hay que hacer un estudio lingüístico porque observamos la estructura del texto:

- una estructura sintáctica
- y una estructura semántica.

Ambas hay que comprenderlas, ambas hay que plasmarlas en el resumen. Para saber hacerlas los pasos son:

a.– Fragmentar el texto:

Descomponer el texto en unidades pequeñas y eliminar las que no nos valen. Pasamos después a la segunda fase.

b.– Identificar las frases o palabras más significativas del texto

c.– Seguir las reglas del centro

VAN DISK dice que hay dos tipos de reglas:

c.1.– Anulación

Selección

Omisión

c.2.– Sustitución

Generalizar

Construir.

c.1.– Anulación: Omitir toda la información no relevante y seleccionar las partes más importante. Eliminar la redundancia (principio de entropía). Haciendo esto eliminas la mitad de la información.

c.2.– Sustitución: Generalizar el texto, que implica el proceso de abstracción. Y construir el informe nuevo.

5ª fase: SÍNTESIS

Es muy difícil aquí aplicar normas. Aquí a partir del análisis se hace una información nueva. Es hacer propiamente el resumen.

Si ya redactamos el resumen, no podemos dar aquí ningún mensaje que no este en el documento original. No podemos introducir ningún elemento semántico que no contenga el documento original. Todo resumen debe seguir estos criterios:

- Mantener la fidelidad al original
- Atenerse al principio de precisión (no redundancias ni repetición)
- Claridad
- Y mayor profundidad posible.
- Entropía
- Objetividad
- Brevedad
- Pertinencia
- Poco distanciamiento cronológico desde la aparición del original hasta la aparición del resumen.

TIPOS DE RESÚMENES:

1.- SEGÚN SU CONTENIDO:

INDICATIVO:

Es el más corto, es el más denso y el más profundo, que no tenga más de 50 palabras. Su función es de alerta al usuario. Se suele emplear en bibliografías, manuales, etc.... Es el más sencillo de hacer.

INFORMATIVOS

Es de menor densidad y de menor profundidad. Describe todos los aspectos relevantes del documento. Aquí debe contener:

- Los objetivos
- El método
- Los resultados
- Las conclusiones.

Este tipo es el máspreciado por el usuario. Tendrá entre 100 y 300 palabras.

INFORMATIVO-INDICATIVO

Las partes centrales son representadas siguiendo el informativo, y las partes más centrales se sigue el indicativo (las partes menos importantes).

ANALÍTICO

Se realiza con menos profundidad y densidad pero es más largo, es decir, mayor representación textual. Entre 500 y 700 palabras. Es poco usual.

CRÍTICO

Esto quiere decir que hace una interpretación del documento, hay muchos autores que no lo reconocen como resumen porque dicen que no es objetivo. Pero otros si que lo incluyen, es decir, sería más como una valoración.

2.- SEGÚN SU AUTORIA

RESUMEN DE AUTOR:

Ventajas: no hay problemas con los derechos de autor, es un avance selectivo de información, es el autor el que mejor conoce el tema, supone economía de tiempo y esfuerzo.

Inconvenientes: el autor no es experto en técnicas de resumir, es incapaz de no plasmar su opinión, es decir, no es objetivo.

Cuando el resumen lo realiza un experto en la materia del documento: hay menos problemas de subjetividad, son más objetivos, sin embargo no son expertos en el arte de resumir.

Cuando lo realizan profesionales de la documentación; son los más capaces en resumir si tienen dominio en el tema.

FUNCIONES DEL RESUMEN:

- Anticipo del original e incluso sustituto del original.
- Actualización informativa
- Supera las barreras lingüísticas
- Facilita la indización.
- Facilita la búsqueda automatizada.