

Lenguajes documentales

Blanca Estela SÁNCHEZ LUNA¹

Resumen

El trabajo explica los llamados *lenguajes documentales* a partir del conocimiento expresado por el ser humano a través del lenguaje escrito y que generalmente se representa en los documentos. En este sentido, la aportación también resalta la importancia de representar el contenido de los documentos a través de un lenguaje controlado con el fin de recuperar fácilmente la información a partir del tema.

Palabras clave: lenguajes controlados, organización del conocimiento.

Abstract

This paper describes the so-called *controlled vocabularies* since a reflection of knowledge that is expressed by human being through written language and that is usually represented in documents. In this sense, this paper also underlines the importance of representing the content of documents through a controlled language in order to retrieve information from the subject.

Keys words: controlled vocabularies, knowledge organization.

Introducción

A lo largo del tiempo, la humanidad ha ido acumulando saber que hoy llamamos conocimiento, este conocimiento se entiende como un proceso mental, inteligente, que origina varias formas de pensamiento, y que en un momento dado, establece una casi-sinonimia con ciencia y con cultura. Los filósofos occidentales suponen que se adquiere conocimiento por las sensaciones o percepciones que se reciben del exterior y que son interpretadas por cada individuo, según su propia idiosincrasia si se añade un

¹ Universidad Nacional Autónoma de México, <bestela@dgh.unam.mx>

componente de racionalidad se llegan a establecer relaciones entre percepción y lenguaje como forma de expresión y de comunicación.

A partir de aquí, se alcanza una estrecha relación entre conocimiento y lenguaje. En este sentido, dicha relación adquiere gran importancia para la documentación.

Conocimiento-Lenguaje-Documento

Los seres humanos comunican las representaciones de su conocimiento por vía oral, escrita o gráfica, por medio del uso de lenguajes orales, escritos y gráficos llamados lenguajes naturales y artificiales.²

En este sentido, las ciencias de la documentación se han especializado en los lenguajes artificiales, conocidos también como vocabularios controlados o lenguajes documentales, como medio eficaz para la recuperación de información.

De aquí que el propósito del presente trabajo sea realizar un recorrido por el tema de los lenguajes documentales a partir del análisis documental como tarea primordial en la organización del conocimiento.

Análisis documental y organización del conocimiento

Desde los primeros tiempos, el ser humano realiza procesos mentales conducentes a la ordenación de todo aquello que compone su vida diaria. Esa ordenación en principio fue inconsciente y gradualmente ha pasado a convertirse en clasificación consciente. Dentro de la bibliotecología el área que se dedica a ordenar el conocimiento es la de lenguajes documentales que corresponde al análisis documental.

El análisis documental es una de las actividades fundamentales que se realiza en cualquier unidad de información, ya que consiste en la representación exacta de los documentos originales, cuya finalidad es que el sistema de información proporcione solamente aquellos documentos que sean relevantes de acuerdo con las necesidades informativas del usuario y de esta manera, hacerlos más fácilmente accesibles a éste.

El análisis documental se divide en tres etapas:

1. Catalogación
2. Indización
3. Resumen

Tiene como finalidad:

- a) Identificar y seleccionar los elementos esenciales de los documentos
- b) Expresar de forma eficaz los puntos de acceso para recuperar la información que contienen los documentos.

² E. Corrás, *Ontologías, taxonomías y tesauros: manual de construcción y uso*. Gijón, Trea, 2005, p. 20.

Actualmente se habla de organización del conocimiento industrial, empresarial, social, científico, documental, bibliotecario, informático, es decir, de una organización que se aplica a todas las tareas de la vida cotidiana.

Por lo tanto, la organización del conocimiento deberá estar en consonancia con las necesidades de la sociedad en un momento dado.

Desde tiempos remotos, han existido bibliotecas con una orientación determinada, por citar algunas:

- Biblioteca de Asurbanipal en Nínave (825 a. C. aprox.)
- Biblioteca de Alejandría por Calímaco (250-240 a. C.) Estableció una ordenación por materias y otra por autores —este paradigma de clasificación sigue vigente.

En la actualidad, entre los sistemas de clasificación más importantes destacan la Clasificación Decimal Dewey, convertida por Henri La Fontaine y Paul Otlet en Clasificación Decimal Universal, así como la Clasificación de la Biblioteca del Congreso de Washington.

Posteriormente, surgen nuevas formas de pensamiento, la informática aparece con sus recursos electrónicos, sistemas expertos, hojas web, metadatos, etcétera, por lo tanto, los conceptos se amplían conforme surgen estas nuevas formas de pensamiento y aparecen nuevas disciplinas que después se entrelazan unas con otras y el conocimiento se vuelve interdisciplinario, transdisciplinario, e intercultural, basado en sistemas nuevos para ser tratados electrónicamente.

Dentro de la nueva terminología podemos citar conceptos en documentación, de los cuales nos ocuparemos de algunos más adelante:

- a) Dominios (de conocimiento).
- b) Ingeniería (del conocimiento).
- c) Minería (de conocimiento).
- d) Webs semánticas.
- e) Ontologías.
- f) Taxonomías.
- g) Tesoros.³

Ahora bien, la causa de la aparición de estos nuevos conceptos es la intrusión de la tecnología informática en el campo de la documentación, por lo que se estructuran programas para computadora cuya finalidad es organizar sistemáticamente el conocimiento a partir de la terminología y del campo semántico de los conceptos.

³ *Ibid.*, p. 32.

Esbozo histórico acerca de los lenguajes documentales

Las primeras manifestaciones de los lenguajes documentales datan de finales del siglo XIX, cuando aparecen las clasificaciones bibliográficas. Estas clasificaciones, inspiradas en la lógica y en los sistemas filosóficos del conocimiento, se fundan en el principio de preordenación y son de carácter enciclopédico.

El concepto moderno de lenguaje documental se consolidó en el siglo XX cuando Cutler introdujo el lenguaje de encabezamientos de materia, basado en los principios de especificidad y de entrada directa, que marca el comienzo del desarrollo de lenguajes documentales especializados.

Los lenguajes especializados nacieron como respuesta a la dispersión del conocimiento, que obligó a una creciente especialización.

Conceptos

En términos generales podemos definir un lenguaje documental como un sistema lingüístico, constituido por términos relacionados entre sí, de acuerdo a normas preestablecidas con respecto al campo semántico de cada término que lo compone, teniendo en cuenta tipo de usuario y entorno disciplinario.

Existen diversas acepciones de lenguaje documental, entre ellas figura la del glosario de la ALA que lo define como un sistema de indización de vocabulario limitado; sistema de indización en el que al asignar descriptores a los documentos, el indizador se limita a una lista específica de términos.

Son "lenguajes convencionales utilizados por una unidad de información para describir el contenido de los documentos para almacenarlos y recuperar las informaciones".⁴

Yves Courcier citado por Gil Urdiciain, los define como:

Unos lenguajes artificiales que permiten generar la representación formalizada de los documentos y de las preguntas que interesan a un grupo de usuarios a fin de recuperar los documentos que respondan a las preguntas.

Todo sistema artificial de signos normalizados, que facilitan la representación formalizada del contenido de los documentos para permitir la recuperación, manual o automática, de información solicitada por los usuarios.⁵

En esta definición se introducen dos elementos (sistema de signos y usuarios) de radical importancia en nuestro estudio. En efecto, uno de los problemas principales en la representación de la información que contiene un documento es el que se refiere al uso de los conceptos que denominar los objetos y fenómenos descritos en el

⁴ C. Guichat y Michel Meou. *Introducción general a las ciencias y técnicas de la información y de la documentación*. París, UNESCO, 1981, p. 21.

⁵ B. Gil Urdiciain, *Manual de lenguajes documentales*. Madrid, Noesis, 1996, p. 18.

documento. En esta operación de representación surgen diversos elementos que deben ser tomados en cuenta a la hora de describir los contenidos documentales. Para Rendón son tres:

1. Un elemento de naturaleza lingüística (término).
2. Un elemento ontológico (el objeto denotado).
3. Elemento lógico-epistemológico (el concepto como forma de pensamiento).

Cada vez que representemos un objeto, necesariamente se debe analizar su definición en el ámbito del sistema lingüístico, del entorno disciplinario y del entorno de los usuarios o del sistema de información.⁶

La aplicación de los lenguajes controlados implica una mayor responsabilidad, a diferencia de la indexación con lenguajes naturales, donde se extraen los términos en forma automática en general y sin la participación del profesional de la información. El documentalista que indiza, después de la interpretación del texto, además de realizar el análisis de cada obra para detectar los conceptos que se discuten en ella, debe ubicarla en el contexto del sistema de información. Sin embargo, en términos generales ambos tipos de sistemas persiguen un mismo objetivo final: la recuperación de la información.

Finalmente, otra definición es: "lenguaje en el que los elementos —palabras— que lo componen, se estructuran de acuerdo con unas reglas preestablecidas, o bien, como sistema lingüístico, en el que las unidades que lo constituyen, se estructuran de acuerdo con unas normas prefijadas". Otra definición es: "lenguaje controlado, al que se le aplican unas normas lógicamente prefijadas y cuyas unidades lingüísticas son términos".⁷

Ahora bien, para establecer un control de términos de un lenguaje documental, se ha de pensar en un proceso previo —*a priori*— y un proceso posterior —*a posteriori*. Este control supone establecer relaciones jerárquicas, asociativas y de equivalencia. Habrá, asimismo, que tener en cuenta que los términos pueden ser simples o compuestos y que estos últimos se podrán formar coordinando sus elementos precisamente *a priori* o *a posteriori*, conocido también como:

- Lenguajes precoordinados.
- Lenguajes poscoordinados.⁸

Características del lenguaje documental

El lenguaje documental es considerado como objeto de estudio de la lingüística documental, cuyas bases teóricas proceden de ramas del conocimiento tales como la lin-

⁶ C. Navarro Peña, *Los lenguajes documentales y su aplicación en la información impresa, digital y multimedia*. México, UNAM/CUB, 2007, p. 99.

⁷ E. Carrís, *op. cit.*, p. 88.

⁸ *Ibid.*, p. 90.

güística general, la semántica, la sintaxis, la terminología y la lexicología. No obstante, el lenguaje documental es anterior a la aparición de la lingüística documental como disciplina.

El lenguaje documental tiene conexiones con diversas ramas del conocimiento: con la inteligencia artificial, la lógica —y, en la medida en que el proceso documental es un proceso de comunicación, y el lenguaje documental es mediador en ese proceso— se le pueden atribuir relaciones con la teoría de la comunicación y con la teoría de la información. Es evidente, por otra parte, la existencia de influencias de las especialidades antes mencionadas (semántica, sintaxis, lingüística) con el lenguaje documental.

A finales de los años sesentas surge la lingüística textual, también denominada ciencia del texto o análisis del discurso que supone un cambio total en las concepciones del análisis del lenguaje en su aspecto semántico. Sus precursores, Perón y Van Dijk, centran la atención en el análisis global del texto a diferencia de las gramáticas precedentes dedicadas al estudio de unidades menores o unidades significativas organizadas bajo determinada estructura.⁹

En un sistema de información el control terminológico se logra a través de la aplicación de un vocabulario controlado que se rige por determinadas normas.

Tipología de los lenguajes documentales

Aunque es más frecuente la denominación "lenguajes documentales", son diversas las denominaciones de los lenguajes cuya función es la indización para localizar información. Algunos autores utilizan términos como: lenguajes de indización, lenguajes terminológicos, lenguajes controlados, lenguajes normalizados, lenguaje documentario, lenguaje de búsqueda de información.¹⁰

En relación con el análisis formal, el lenguaje documental completa el proceso de catalogación dotando al soporte de la descripción de puntos de acceso temáticos.

De acuerdo con la tipología establecida por Gil Urdujain vale la pena transcribir los diversos criterios de tipificación de los lenguajes documentales que la autora señala, los más generalizados son: el de control, el de continuación de los términos y el estructural.

Dependiendo del control ejercido sobre el vocabulario, los lenguajes se organizan en dos categorías: libres y controlados.

Control	Libres	Listas descriptores libres
	Controlados	Clasificaciones, tesauros, catálogos de autoridad

Tabla 1. Categorías de los lenguajes.

⁹ B. Gil Urdujain, *op. cit.*, p. 18.

¹⁰ E. Currás, *op. cit.*, p. 14.

Los lenguajes libres, cuyo principio es la poscoordinación, se componen de un vocabulario no predefinido que se va generando a partir de la realización de procesos de indización (listas de descriptores libres y palabras clave).

El lenguaje libre se caracteriza por una ambigüedad semántica muy grande, en el cada concepto puede expresarse por una serie de sinónimos y de perífrasis, y cada palabra puede designar varios conceptos (polisemia).

Las listas de descriptores libres presentan un orden alfabético para cada término, y son independientes unos de otros, por lo que la búsqueda de documentos se multiplicará por cada concepto que el usuario formule, sean estos términos sinónimos o polisémicos y cuya finalidad será recuperar la mayor cantidad de documentos que sea posible. Aunque el lenguaje libre presenta dificultades en el momento de la recuperación de información, aventaja al controlado en lo que se refiere a su utilización en los procesos de indización, al no necesitar, como aquél, traducir los conceptos en descriptores de un tesoro o en los encabezamientos de materia conformados en una lista.

La sistematización de los lenguajes documentales, según el criterio de coordinación, se realiza en función del momento en que se combinan los elementos que los componen. Si los términos se combinan en el momento de la descripción, el lenguaje será precoordinado, y si se hace en el momento de la recuperación, se tratará de un lenguaje poscoordinado.

Coordinación	Precoordinados	Clasificaciones Listas de encabezamientos de materia
	Poscoordinados	Listas de descriptores libres Listas de palabras clave Tesaurus

Tabla 2. Tipos de lenguajes de acuerdo al criterio de coordinación.

En los sistemas precoordinados los términos están unidos de acuerdo con unas reglas sintácticas establecidas por el propio lenguaje. En los sistemas poscoordinados los términos se combinan de acuerdo con la lógica de cada usuario, aunque también subyacen principios generales y universales de lógica booleana y teoría de conjuntos.

En cuanto a la estructura, existen tres tipos: jerárquica, combinatoria y sintáctica. Los lenguajes de estructura jerárquica organizan sistemáticamente el conocimiento, es decir, agrupan las materias en áreas categoriales. A la estructura jerárquica responden determinadas clasificaciones en las que la dependencia de un concepto con respecto a otros se refleja por medio de códigos alfabéticos, numéricos o alfanuméricos que facilitan su localización en la estructura arborescente.

Los lenguajes de estructura combinatoria permiten la libre combinación de los términos que los componen entre sí, según las necesidades de indización. Los léxicos documentales y los tesaurus corresponden a esta categoría.

A la estructura sintáctica pertenecen los lenguajes que recurren a una sintaxis mediante la cual se pueden representar y poner en relación, los contenidos de los documentos.

Presentados de forma esquemática, los lenguajes documentales, en función de su estructura, de acuerdo a Gil,¹¹ pueden ser:

Estructura:	Jerárquica	clasificaciones jerárquicas
	Combinatoria	léxicos documentales - tesoros
	Sintáctica	de gramática simple de gramática elaborada

Tabla 3. Lenguajes documentales en función de su estructura.

Definición de los distintos lenguajes

Listas de palabras clave

Listas no estructuradas, salvo su ordenación alfabética, de conceptos que han sido seleccionados por métodos automáticos para describir el contenido de los documentos analizados. El método de determinación de las listas es singular en cada documento de modo que no permiten la descripción de otros materiales.

Quienes apoyan la utilización del lenguaje libre señalan la ventaja de su valor informativo pues afirman que se trata, a menudo, de términos muy específicos y actuales que permiten seguir de cerca la evolución terminológica de un tema determinado. A esta ventaja añaden otras como el ahorro que supone no tener que elaborar y mantener un lenguaje documental. Sus detractores señalan como principal desventaja la ausencia de control eficaz a causa de la ambigüedad del lenguaje natural. El usuario está sometido a las variaciones semánticas y ortográficas de un mismo término y a las expresiones sinónimas de un mismo concepto.¹²

Listas de descriptores libres

El lenguaje natural de partida contiene ambigüedades, redundancias y conceptos implícitos que hacen difícil manejar sus expresiones. El manejo de las ambigüedades es difícil ya que una palabra puede tener varios significados (homonimia y polisemia léxicas) así como dificultades sintácticas (ambigüedad estructural). Sin embargo, este tipo de listas se ha hecho posible gracias a las computadoras.¹³

¹¹ B. Gil Urdiciain, *op. cit.*, p. 22.

¹² *Ibid.*, p. 26.

¹³ J. A. Moreira González, *El contenido de los documentos textuales: su análisis y representación mediante el lenguaje natural*. Gijón, Trea, 2004, p. 145.

Por otra parte, el reconocimiento del hecho de que los encabezamientos de materia se prestan mal a la recuperación automática llevó a Tabor a sustituir el encabezamiento, compuesto generalmente por elementos unitarios no relacionados entre ellos, por el sistema unitérmico, basado en el principio de poscoordinación, que permite la elección del término de indexación en respuesta a necesidades de información reales. Tanto el lenguaje de unitérmicos como el de descriptores —términos compuestos definidos por Moores como *uniconceptos*—, están compuestos por listas de conceptos sin más estructuración que la organización alfabética bajo la que se presentan. Son el resultado del análisis de documentos que permitirán la descripción posterior de otros materiales. Las listas de descriptores libres no limitan la incorporación de nuevos conceptos.¹⁴

Clasificaciones

La mayoría de los sistemas de clasificación que existen intentan representar de una manera lógica el conocimiento humano. Existen sistemas que usan notaciones numéricas o alfanuméricas para lograr su objetivo.

Clasificar, en términos generales, es el acto de organizar el universo del conocimiento en algún orden sistemático. Ha sido considerada la actividad más fundamental de la mente humana. El acto de clasificar consiste en el dicotómico proceso de distinguir cosas u objetos que poseen cierta característica de aquellos que no la tienen y agrupar en una clase cosas u objetos que tienen la propiedad o características común.¹⁵

La mayoría de las clasificaciones bibliográficas están inspiradas en los principios de clasificación de la lógica y de los sistemas filosóficos del conocimiento. Las clasificaciones son los lenguajes documentales de mayor tradición y antigüedad.

Tres de los sistemas más conocidos y usados en el mundo son, como ya se mencionó: la Clasificación Decimal de Dewey, la Clasificación de la Biblioteca del Congreso de Washington y la Clasificación Decimal Universal.

Presentan una distribución sistemática de conceptos en diversas categorías o clases, de tal manera que cada materia tiene un lugar predefinido. Su utilización es habitual en bibliotecas por su carácter enciclopédico. Se le reconoce como, inconscientemente, la falta de operatividad por la rigidez de su estructura.¹⁶

¹⁴ B. Gill Underlain, *op. cit.*, p. 26.

¹⁵ L. M. Chan, *Immeroff's Guide to the Library of Congress Classification*, Littleton, Libraries Unlimited, 1980, p. 200.

¹⁶ B. Gill Underlain, *op. cit.*, p. 26.

Listas de encabezamientos de materia

Constituyen un tipo de lenguaje documental controlado y están compuestas por términos sueltos o combinados entre sí para formar conceptos o frases que representan el contenido de una obra. Castañón, en su obra *Los encabezamientos de materia*, define a éstos como la palabra o palabras con que se encabeza, inicia, o principia la relación descriptiva de un libro en una ficha catalográfica. También señala que son a su vez el elemento base del ordenamiento, es decir, son las palabras que se toman en cuenta para alfabeticar las fichas en el catálogo. En este catálogo un libro puede ser representado varias veces a través de diferentes encabezamientos de autor, encabezamientos de título y encabezamientos de materia.¹⁷

La definición anterior alude al catálogo tradicional en tarjetas que imperó hasta aproximadamente la década de los noventa, sin embargo, la obra de la maestra Castañón constituye aún la base teórica que sirve de guía para la conformación y actualización de listas de encabezamientos para bibliotecas o para los llamados *catálogos de autoridad de materia* que resultan, hoy día, de mucha utilidad a los usuarios en la recuperación de información dada su interacción con los catálogos bibliográficos.

Los encabezamientos de materia se componen por una serie de elementos básicos conocidos como referencias, que en términos generales controlan la sinonimia, polisemia y homonimia así como el campo semántico del término o concepto mediante los términos relacionados, específicos y genéricos.

Las referencias son elementos fundamentales en los catálogos de materia y de autor puesto que en ellas descansa la validez del encabezamiento de materia único.

Existen varios tipos de referencias, a saber:

1. Referencias de *véase*.
2. Referencias de *véase también*.
3. Referencias globales.
4. Notas de alcance.

Las referencias de *véase* sirven para controlar sinónimos, formas variantes de escritura o formas diferentes de expresar el mismo concepto de acuerdo a la localidad geográfica. Por su parte, las referencias de *véase también* son los términos relacionados, genéricos y específicos con relación al encabezamiento de materia que se ha establecido como aceptado. Finalmente, las *referencias globales* y *notas de alcance* son textos explicativos que ayudan a esclarecer el uso apropiado del encabezamiento único o aceptado.

Ahora bien, de acuerdo a la forma o tipología de los encabezamientos de materia, éstos se dividen en:

¹⁷ B. Castañón Moreno, *Los encabezamientos de materia*. México, UNAM, 1974, p. 2.

- Encabezamientos de materia formados por un sustantivo (Animales).
- Encabezamientos formados por un sustantivo y un adjetivo (Animales salvajes).
- Encabezamientos formados por dos sustantivos unidos por una preposición (Guerras en la Biblia).
- Encabezamientos de materia formados por dos o más palabras unidas por una conjunción (Guerra y civilización).
- Encabezamientos formados por locuciones, frases hechas o convencionales (Teoría del conocimiento).

Y de acuerdo con los signos que intervienen en la estructura de los encabezamientos de materia éstos pueden ser:

- Encabezamientos de materia invertidos (actualmente en desuso, pues su utilidad recae en los catálogos de tarjetas).
- Encabezamientos de materia con aclaración (sirven para aclarar confusiones entre términos).
- Encabezamientos de materia subdivididos.¹⁸

Entre las ventajas que tienen los encabezamientos de materia se encuentran que el usuario no tendrá que imaginar todos los sinónimos posibles del tema que busca y que permite conocer las relaciones entre términos.

Por regla general las listas son monolingües y se componen de encabezamientos y subencabezamientos.

Los subencabezamientos se caracterizan por representar materias en cuya simbolización se emplea el guión de separación entre dos palabras y sirven para delimitar el tema principal, por ejemplo, México —Historia— Siglo XX, es decir, una obra que trata sobre la historia de México en el siglo XX específicamente.

Por su naturaleza, los subencabezamientos se dividen en:

- a) Subencabezamientos de forma (Manuales, directorios)
- b) Subencabezamientos geográficos (México, América).
- c) Subencabezamientos cronológicos (Siglo XX, Edad Media).
- d) Subencabezamientos de materia o asunto (Historia, Aspectos Sociales)¹⁹

Tesaurus

Es un lenguaje documental que recoge palabras simples o compuestas que integran las expresiones utilizadas por los autores de los documentos de un área especializada que acreditan la garantía literaria de los términos; es un lenguaje de indexación que permite

¹⁸ *Ibid.*, p. 111.

¹⁹ *Ibid.*, p. 16.

representar los temas de los documentos y tiene la capacidad para indizar los documentos de una biblioteca o de una base de datos especializada o para marcar términos de recuperación en una biblioteca digital. También es un lenguaje controlado que los términos que incluye están validados por un proceso en el que se establecen relaciones entre ellos para mantener la univocidad de sus descriptores y se determina la preferencia de un término sobre otros posibles.²⁰

La ISO 2788/TC46 define la palabra *tesoro* desde el punto de vista funcional y estructural. Funcionalmente es un instrumento para controlar la terminología al trasladar a un lenguaje más estricto la lengua natural de los documentos. Estructuralmente es vocabulario controlado y dinámico de términos compuesto por descriptores relacionados entre sí semánticamente (jerarquía, asociación, equivalencia) y que se aplica a un dominio particular del conocimiento.

Lenguajes de estructura sintáctica

Constituidos por términos susceptibles de conectarse entre sí con base en una sintaxis más o menos compleja que determina la subdivisión de estos lenguajes sintácticos simples o elaborados.

Los lenguajes de estas características se fundamentan en el establecimiento de unas reglas sintácticas más o menos complicadas, que van desde la creación de indicadores de rol, que representan el papel sintáctico de cada descriptor en la frase documental, hasta el establecimiento de sistemas de nexos, que asocian los diferentes conceptos de que consta la frase según la relación que queramos establecer entre los descriptores.

Los lenguajes de estructura sintáctica aparecieron en un momento en que no había herramientas para el almacenamiento masivo, por lo que con el paso del tiempo cayeron en desuso.²¹

Taxonomías

Es otro concepto surgido a partir de la irrupción de la informática en la documentación. Se define como el campo de la ciencia que clasifica la vida (la vida comprende, además de los seres vivos, toda idea y actividad humana). Asimismo, se define como un lenguaje controlado. Una lista organizada de palabras y frases, o sistemas de notación que se usa para iniciar un proceso de indización y recuperación de información.²²

²⁰ C. Naumis Peña, *op. cit.*, p. 121.

²¹ B. Gil Urdiales, *op. cit.*, p. 27.

²² E. Carrás, *op. cit.*, p. 60.

Ontologías

Actualmente se habla de las ontologías como sinónimo de lenguajes documentales, pero con una estructura muy diferente a la de los lenguajes documentales, Márquez Solís, en su obra *La Web semántica*, nos dice al respecto:

- Las ontologías proveen de una comprensión compartida y consensuada del conocimiento de un dominio que puede ser comunicada entre personas y sistemas heterogéneos. Fueron desarrolladas en el área de Inteligencia Artificial (IA) para facilitar el intercambio y reutilización del conocimiento.
- El término ontología se remonta a la civilización griega, Aristóteles lo definió como la ciencia del ser, aunque en informática se refiere a un vocabulario compartido que describe un determinado dominio y que se define en términos de un lenguaje formal de manera que sea manipulable automáticamente.
- Generalmente las ontologías se representan mediante clases, propiedades, atributos de las clases, relaciones entre clases y restricciones que pueden producirse tanto en los atributos como en las propiedades.²³

A continuación se muestra un ejemplo de ontología para definir: fuentes naturales de agua.

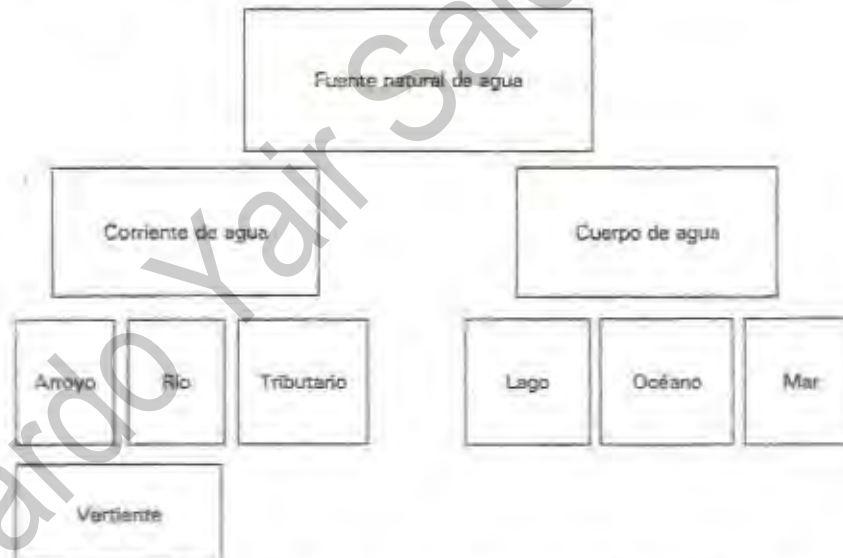


Figura 1. Ontología para definir fuentes de agua. Fuente: S. Márquez Solís, *op. cit.*, p. 29.

²³ Santiago Márquez Solís, *La Web semántica*, lulu.com, 2007, p. 28.

Lars Marius Garshal, citado por Carrás, nos dice, en términos generales, que las ontologías dentro de las ciencias de la computación son modelos para describir el mundo, y consisten en conjuntar temas (bases de conocimiento), propiedades y tipos de relaciones.

La diferencia entre un tesoro y una ontología estriba en la estructura. Mientras que en el tesoro se parte de una ordenación de los términos que lo componen en jerarquías y relaciones semánticas y sintácticas, en las ontologías se concibe la ordenación de una manera diferente teniendo en cuenta ciertas peculiaridades y propiedades de los términos así como de las categorías.²⁴

Web semántica

La web semántica es un concepto acuñado por Tim Berners-Lee quien la define de las siguientes maneras:

- El primer paso es colocar los datos en la web de un modo en que las máquinas puedan entenderlos naturalmente o convertirlos a esa forma. Esto crea lo que yo llamo web semántica: una red de datos que pueden ser procesados directa o indirectamente por máquinas.²⁵
- La web semántica es una extensión de la web en la cual la información se da mediante un significado bien definido, lo que facilita que los ordenadores y la gente trabajen en cooperación.²⁶

La web semántica se articula a partir de conceptos como el enlazado de información, el hipertexto y los sistemas jerárquicos. Tim Berners-Lee lo esquematizó como se muestra en la figura 2.

La web semántica mantiene los principios que han hecho un éxito de la web actual, como son los principios de descentralización, compartición, compatibilidad, máxima facilidad de acceso y contribución, o la apertura al crecimiento y uso no previstos de antemano. En este contexto un problema clave es alcanzar un entendimiento entre las partes que han de intervenir en la construcción y explotación de la web: usuarios, desarrolladores y programas de muy diverso perfil. La web semántica resalta la noción de ontología del campo de la Inteligencia Artificial como vehículo para cumplir este objetivo.²⁷

²⁴ Lars Marius Garshal, apud E. Carrás, op. cit., p. 36.

²⁵ *Weaving the web*, 1999, apud S. Márquez Solís, p. 24.

²⁶ *The Semantic Web*, *Scientific American*, mayo de 2001, *ibid.*, p. 24.

²⁷ Pablo Castellá, *La web semántica* [en línea]. <<http://irantxa.ii.uam.es/~castello/publications/castella-ucjm03.pdf>>. [Consulta: 18 de septiembre, 2009], p. 5.

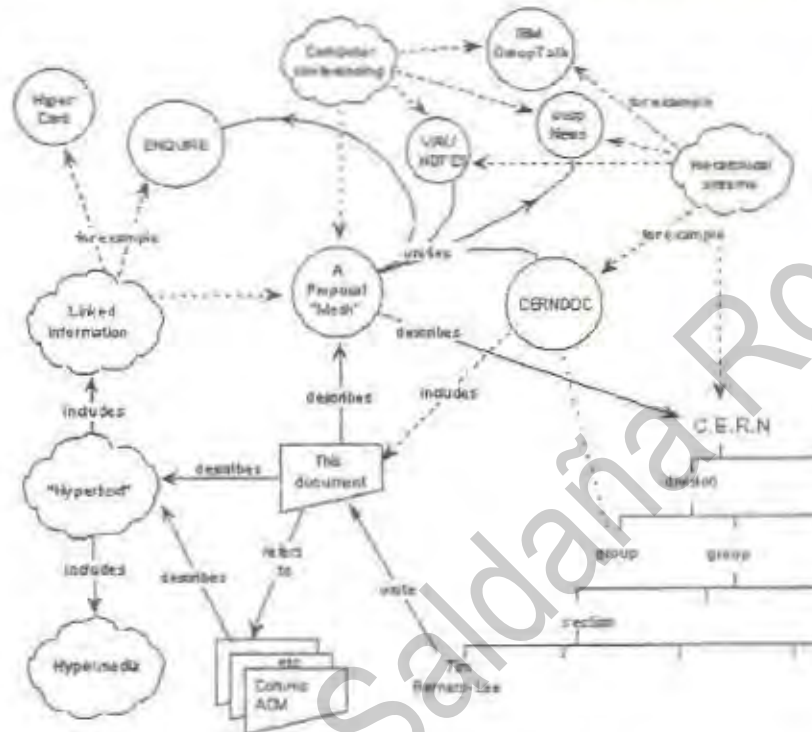


Figura 1. La Web Semántica de Tim Berners-Lee.

Crterias y condiciones para una buena indización

Si se parte del hecho que la indización tiene como finalidad extraer uno o más conceptos que representen el contenido temático del documento, existen cinco criterios que el indizador debe considerar en la elección de los términos que conformarán su lenguaje documental, independientemente del tipo de lenguaje que elija.

1. **Especificidad:** rechazar descriptores demasiado generales o demasiado particulares respecto a las nociones que expresa el documento. Cuando se consigue sus efectos se notan en la relevancia de la recuperación que consiste en que un documento sea útil para los propósitos que causaron una búsqueda por parte del usuario.
2. **Sistematización:** significa aportar la mayor cantidad posible de información, utilizando la menor cantidad posible de signos. Consiste en expresar la globalidad del contenido del mensaje, utilizando la menor cantidad posible de descriptores. Con la sistematización se pretende evitar la redundancia.