

## COMUNICACIÓN TÉCNICA

### PROYECTO CATÁLOGO – MÓDULO ONTOLOGÍAS

Entre los objetivos generales de la formación del Comunicador Técnico aparece la producción de soluciones integradas de comunicación técnica a través de la gestión e ingeniería de la información. Y entre las competencias específicas figura la generación de clasificaciones y catálogos a partir de jerarquías de clases construidas con herramientas específicas (Protégé).

Trataremos de ordenar algunos conceptos para comprender mejor lo anterior.

No escapa a nadie que la revolución tecnológica de las últimas décadas, las redes e Internet ha cambiado la percepción de la representación y el valor de la información. Sin embargo, muchas veces los volúmenes de información disponible son tan grandes que se hace difícil encontrar una forma clara de abordarla. Esto se debe a una carencia de control y estructura en la información.

El desafío fundamental que se plantea es organizar los contenidos de información para que el usuario encuentre el mejor camino, no el único, hacia el conocimiento y la comprensión de esa información.

Por contenidos se entiende todos aquellos 'trozos' de información u objetos informativos a organizar, estructurar y clasificar: textos, imágenes, videos, etc.

- Una primera tarea es analizar la naturaleza de los contenidos a organizar.
- Luego es necesario determinar su granularidad, que se refiere al nivel de descomposición o grado en que pueden ser divididos estos contenidos.

El nivel mínimo de granularidad de un contenido es aquel grado de descomposición en el que una determinada información sigue manteniendo su significación comunicativa, o el grado en el que físicamente no puede seguir descomponiéndose.

Por ejemplo, habrá casos en los que una misma imagen pueda ser utilizada en contextos diferentes - en asociación con diferentes objetos informativos- o que por el contrario, se encuentre vinculada de forma permanente a un contexto determinado. Este caso podría ser el de una imagen embebida en un documento con formato hermético, como PDF. Este documento sería entonces el nivel mínimo de granularidad: *una unidad de contenido*.

- La siguiente tarea es catalogar o describir dichas unidades de contenido, es decir aplicarles metadatos.

Los metadatos son 'información acerca de información'. *Aplicar metadatos significa caracterizar cada unidad de contenido a través de una serie de pares 'atributo-valor/es'*.

*Los atributos pueden ser formales o administrativos - lenguaje, formato, fecha de creación, tamaño...-, o pueden ser descriptivos – título, resumen, palabras clave... -*. Mientras que los administrativos tienen una función relacionada principalmente con la gestión de los contenidos, los descriptivos tienen por objeto facilitar y mejorar la recuperación de información por parte del usuario.

Ejemplo de metadatos:

Atributo	Valor/es
Fecha	30/08/1998
Título	Evaluación de la calidad de uso
Palabras clave	test de usuarios; evaluación heurística; card sorting
Categoría	Desarrollo web

El número y tipos de metadatos a aplicar dependerán de la naturaleza de los contenidos. Si por ejemplo lo que se está catalogando es un conjunto fotográfico, los metadatos a aplicar podrían ser: autor, formato, tamaño, peso, resolución, color, palabras clave, ...

*De cara al usuario y la recuperación de información, los tipos de metadatos más útiles son los descriptivos, ya que indican explícitamente sobre qué trata la unidad de contenido.*

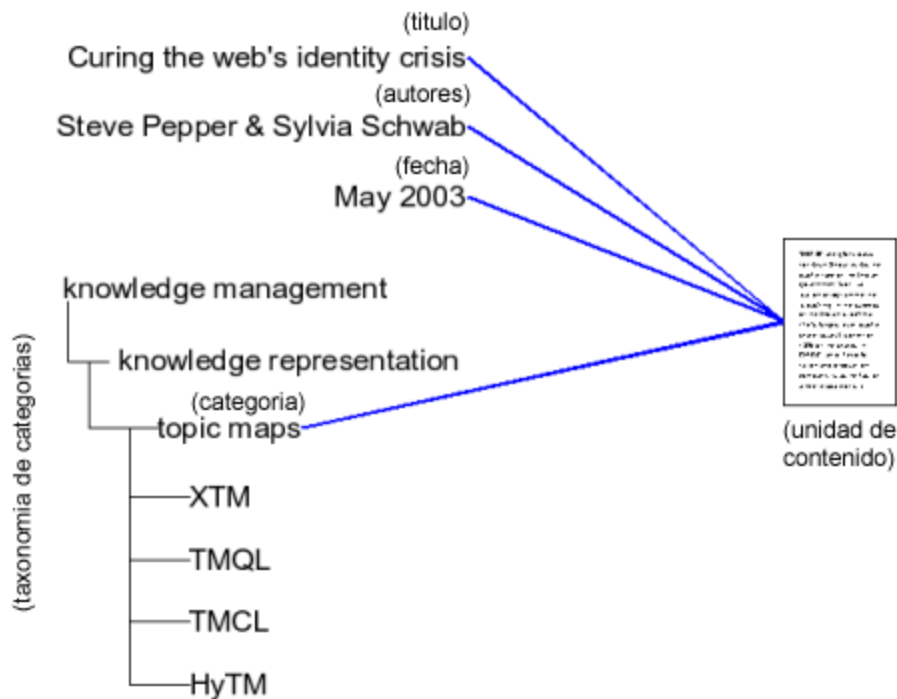


Figura 1. (Garshol, 2004)

En la figura de ejemplo podemos observar el esquema de una unidad de contenido descrita mediante metadatos, donde las líneas azules representan los atributos (título, autores, fecha y categoría) que vinculan cada unidad de contenido con sus correspondientes valores para estos atributos.

### Palabras Clave, Vocabularios Controlados e Indización

Las palabras clave son un conjunto de términos que representan los conceptos más significativos de la unidad de contenido. Resumen su contenido a través de los conceptos más relevantes tratados.

El primer inconveniente que presentan está relacionado directamente con la ambigüedad propia del lenguaje natural: *polisemia* y *sinonimia*. Un término puede representar más de un concepto diferente y un mismo concepto podría ser representado por varios términos diferentes.

Para evitar estos problemas que podrían repercutir en su utilidad descriptiva, es necesario contar con herramientas o mecanismos que aseguren la consistencia de los valores definidos para estos metadatos, que controlen la terminología empleada.

Un vocabulario controlado es una lista o índice de términos que establece relaciones unívocas y precisas entre ellos, así como con los conceptos representados. Dependiendo del número y tipo de relaciones definidas entre los términos se suele hablar de *anillos de sinónimos* o *listas de autoridades* como ejemplos de vocabularios controlados simples, o de *tesauros* como un tipo de vocabulario controlado más complejo. Mientras que en los primeros las relaciones entre términos son únicamente de sinonimia y de preferencia, en los *tesauros* el número de relaciones entre términos es mayor:

*Jerárquicas:*

- TG o Término Genérico: Relación del término con un término "padre", del que desciende.
- TE o Término Específico: Relación del término con un término "hijo", más concreto.

*Equivalencia:*

- USE: Relación entre el término no preferente y el término preferente, o término a usar (más aceptado).
- UP o Usado Por: Relación entre el término preferente y el término no preferente.

*Asociativas:*

- TR o Término Relacionado: Relación entre el término y otro término afín.

*Semánticas:*

- NA o Nota de Alcance: Explicación o descripción del significado del término

**Indizar** comprendería la tarea de catalogar o describir los recursos a través de palabras clave que forman parte de un índice terminológico o vocabulario controlado.

La primera ventaja de usar un vocabulario controlado en la indización de las unidades de contenido tiene que ver con la consistencia de las palabras clave asignadas: cada concepto tendrá una correspondencia unívoca con un término, y viceversa, eliminando así los problemas de polisemia y sinonimia.

Otra ventaja de su uso estará relacionada con la recuperación o búsqueda de estas unidades de contenido por el usuario mediante los sistemas de búsqueda y navegación.

### Categorías, Taxonomías y Clasificación

*La distinción entre palabras clave y categorías – y por tanto entre indización y clasificación – es una distinción formal.* Mientras que las palabras clave asociadas a una unidad de contenido representan sus conceptos más significativos, la categoría asociada describe el tema principal bajo el que se engloba la unidad de contenido.

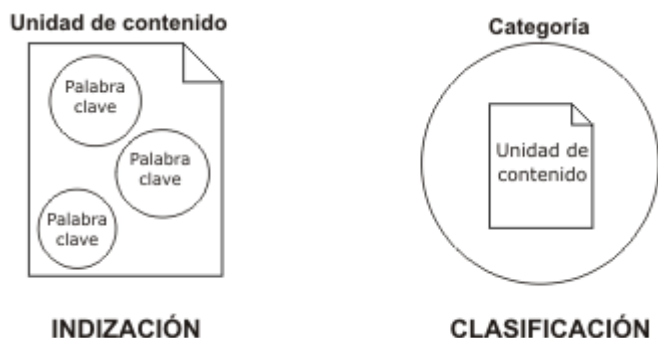


Figura 2: Diferencia entre palabras clave y categorías

Una unidad de contenido podría pertenecer a más de una categoría, aunque esta forma de clasificar haría más difusa, todavía, la diferencia entre categorías y palabras clave.

Las categorías y sus relaciones se organizan en *taxonomías de categorías*.

El término taxonomía proviene del área de la Biología que estudia la clasificación de los seres vivos en estructuras jerárquicas. De forma general podemos decir que una taxonomía es una estructura jerárquica de elementos o grupos de elementos, por lo que es normal encontrar utilizado este término en relación con los tesauros, ontologías, sistemas de clasificación, etc.

En cuanto a la organización de la información, el término taxonomía se entiende como la organización jerárquica del conjunto de categorías bajo las que clasificar las unidades de contenido. Un ejemplo de taxonomía de categorías es el que se observa en la figura 2, donde las líneas negras representan las relaciones jerárquicas entre categorías, y la línea azul el vínculo de pertenencia entre unidad de contenido y categoría.

**Clasificar** comprendería la tarea de catalogar o describir los recursos a través de las categorías de una taxonomía indicando a cuál de ellas pertenecen.

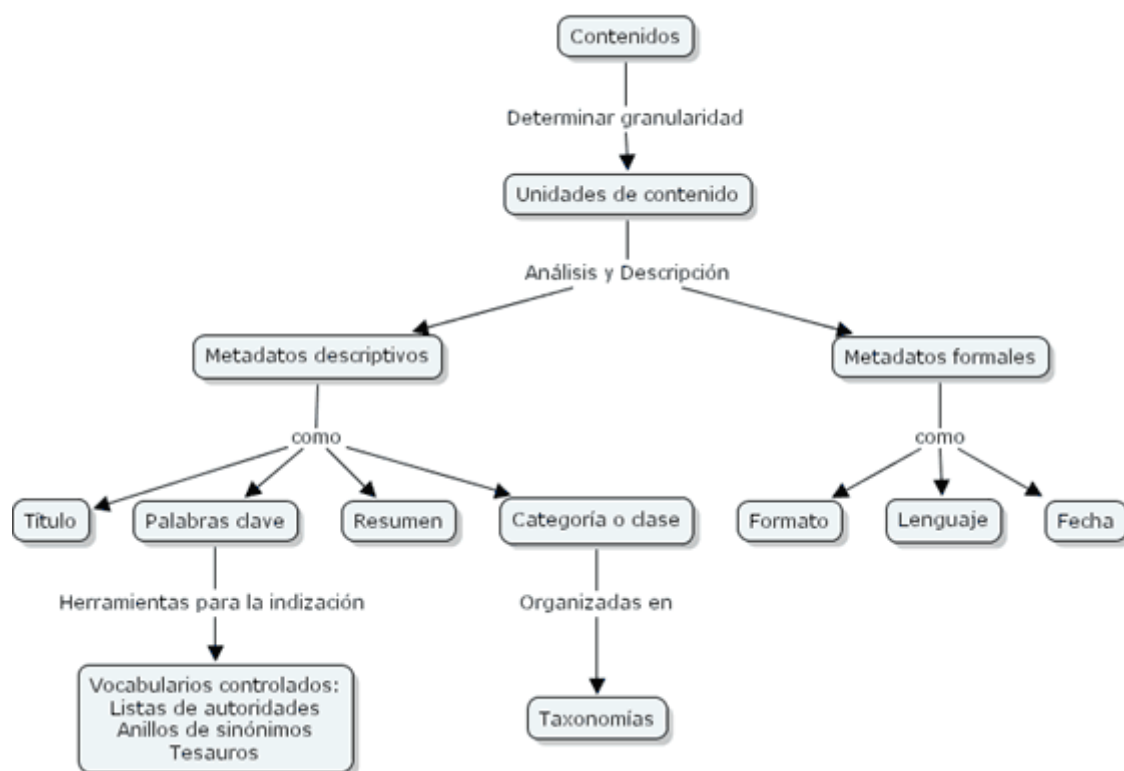


Figura 3: Esquema del proceso de organización de contenidos

### Clasificaciones facetadas

La clasificación facetada parte de la premisa de que una unidad de contenido puede ser descripta a través de varias dimensiones o facetas, cada una de las cuales contiene su propia relación de posibles valores o categorías.

Siguiendo el enfoque de este artículo, todos aquellos tipos de metadatos donde más de una unidad de contenido pudiera compartir valor, pueden ser considerados facetas. No tendría sentido, por ejemplo, considerar facetas a tipos de metadatos como 'título' o 'resumen', ya que sus posibles valores serán únicos para cada unidad de contenido. En cambio, 'categoría temática', 'fecha', 'formato' o 'idioma' sí podrían ser consideradas facetas de una misma clasificación.

La clasificación facetada resulta realmente útil cuando las unidades de contenido son lo suficientemente homogéneas como para que todas puedan ser descriptas a través de un conjunto de metadatos comunes.

Un ejemplo podría ser la organización de anotaciones en una bitácora, como es el caso de efímera.org:

Clasificado en:

- Autor: [Juanjo Seixas](#)
- Espacio: [Galiza](#)
- Forma: [Sitio web](#)
- Tópico: [Diseño](#)
- Tiempo: [Diciembre de 2004](#)

Figura 4: Pie de anotación de efímera.org

En este blog, cada anotación o post se encuentra descripto por un conjunto de metadatos facetados comunes: espacio, forma, tópico y tiempo.

### Anotaciones ordenadas por tópicos

[Arte](#) (44) | [Cartografía](#) (0) | [Cine](#) (7) | [Cómico](#) (1) | [Cuerpo](#) (2) | [Dios](#) (9) | [Diseño](#) (142) | [Educación](#) (18) | [Fotografía](#) (34) | [Ilustración](#) (23) | [La ciudad y yo](#) (58) | [Literatura](#) (5) | [Moda](#) (1) | [Música](#) (12) | [Poesía](#) (7) | [Política](#) (64) | [Sociedad](#) (107) | [Tecnología](#) (37) |

### Anotaciones ordenadas por su forma

[Anotación](#) (29) | [Aplicación](#) (5) | [Artículo](#) (14) | [Blog](#) (15) | [Código](#) (1) | [Cómico](#) (5) | [Correo-e](#) (2) | [Disco](#) (3) | [Exposición](#) (8) | [Flash](#) (2) | [Gente](#) (3) | [Ilustración](#) (7) | [Juego](#) (2) | [Libro](#) (21) | [Manifiesto](#) (3) | [Mapa](#) (8) | [Objeto](#) (1) | [Película](#) (16) | [Producto](#) (3) | [Proyecto](#) (12) | [Revista](#) (3) | [Sitio web](#) (48) | [Tipografía](#) (7) | [Tutorial](#) (1) | [TV](#) (3) |

### Anotaciones ordenadas espacialmente

[África](#) (1) | [Alemania](#) (3) | [Argentina](#) (3) | [Australia](#) (2) | [Austria](#) (1) | [Brasil](#) (1) | [Canadá](#) (1) | [Checoslovaquia](#) (1) | [Chile](#) (2) |

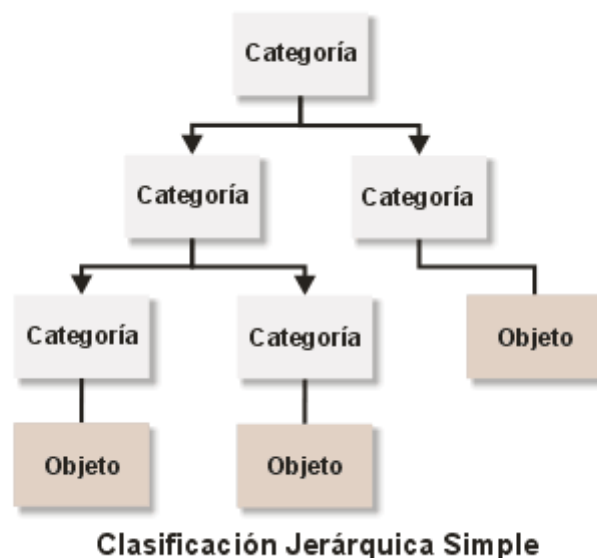
Figura 5: Sección de archivos de efimera.org

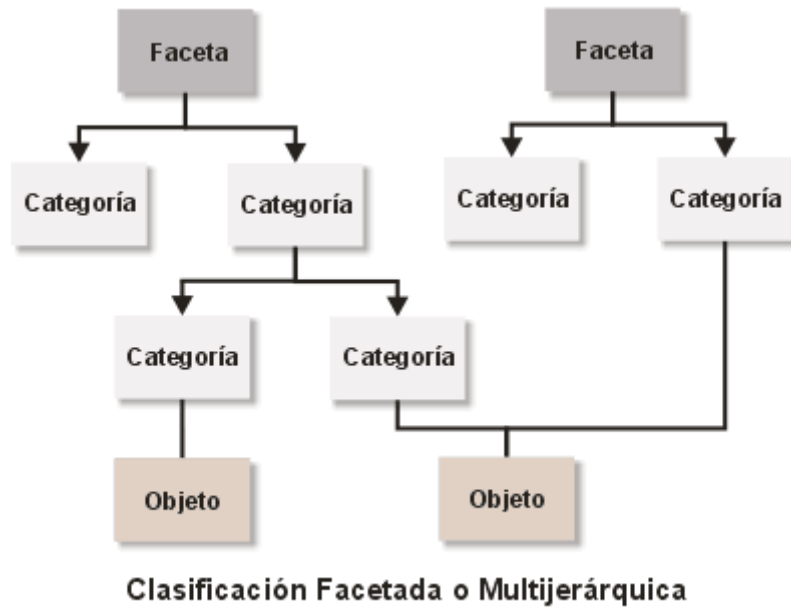
Las clasificaciones facetadas se utilizan para organizar conjuntos de objetos (sitios web, productos, documentos, ...), con una homogeneidad suficiente como para que puedan ser descriptos por un número determinado de atributos o propiedades (facetadas y categorías) y sus valores (pertenencia a categorías), o lo que es lo mismo, por un conjunto de *metadatos facetados*.

### Modelo Facetado vs. Modelo Jerárquico Simple

La mejor forma de entender qué es un Sistema de Clasificación Facetada es mediante su comparación con sistemas de clasificación más comunes, como es el modelo Jerárquico Simple.

En los siguientes esquemas podemos ver las diferencias entre los dos modelos:





La clasificación facetada se diferencia de la clasificación jerárquica simple en que los objetos son caracterizados a través de múltiples dimensiones o facetas, cada una de las cuales posee su propio grupo de categorías.

### Facetas, Categorías y Objetos

Cada faceta está compuesta por un conjunto de categorías. Según cómo sean las relaciones entre las categorías que formen la faceta, podemos diferenciar básicamente entre:

- *Faceta Jerárquica*: Las categorías poseen relaciones de tipo jerárquico (parte-todo, general-específico...)
- *Faceta Plana*: Al margen de la relación (implícita) que les confiere pertenecer a una misma faceta, las categorías no tienen ningún tipo de relación (explícita) entre ellas.
- *Otras*: Las categorías poseen otro tipo de relaciones (explícitas) aparte de las de tipo jerárquico (por ejemplo, relaciones de tipo asociativo).

Las facetas son contenedores mutuamente excluyentes de categorías. Esto quiere decir, que una misma categoría no puede pertenecer a dos facetas diferentes.

No tendría sentido crear diferentes facetas que pudieran contener un mismo tipo de categorías, ya que entonces este tipo de clasificación perdería su función: caracterizar un objeto desde diferentes puntos de vista, para así obtener una descripción global y multidimensional del objeto.

Un objeto, por tanto, es descrito a través de su 'pertenencia', 'ocurrencia' o 'aparición' en las diferentes categorías.

Por la propia definición de categoría, se da por sentado que una categoría puede contener cualquier número de objetos. Esos objetos podrían (y normalmente deberían) aparecer en diversas categorías de diferentes facetas.

En cambio, que un mismo objeto pueda aparecer en diferentes categorías de una misma faceta depende del modelo conceptual con que definamos nuestra clasificación. Esto quiere decir que somos nosotros los que decidimos si las categorías de una misma faceta son mutuamente excluyentes o no, dependiendo de la finalidad con la que creemos nuestra clasificación y la naturaleza de los objetos a clasificar.

### Ejemplo

El siguiente ejemplo, en el que se clasifican 'Establecimientos Hoteleros', muestra las ventajas de las clasificaciones facetadas, y sus diferencias respecto a las jerárquicas simples.

*Modelo Jerárquico simple:*

- Establecimiento Hotelero
  - Hotel de Ciudad
    - Vacacional
      - ◆ Sólo Desayuno
      - ◆ Desayuno, Almuerzo y Cena
    - Convenciones y Eventos
      - ◆ Sólo Desayuno
      - ◆ Desayuno, Almuerzo y Cena
    - Comercial
      - ◆ Sólo Desayuno
      - ◆ Desayuno, Almuerzo y Cena
  - Hotel de Montaña
    - Vacacional
      - ◆ Sólo Desayuno
      - ◆ Desayuno, Almuerzo y Cena
    - Convenciones y Eventos
      - ◆ Sólo Desayuno
      - ◆ Desayuno, Almuerzo y Cena
    - Comercial
      - ◆ Sólo Desayuno
      - ◆ Desayuno, Almuerzo y Cena
  - Hotel de Playa
    - Vacacional
      - ◆ Sólo Desayuno
      - ◆ Desayuno, Almuerzo y Cena
    - Convenciones y Eventos
      - ◆ Sólo Desayuno
      - ◆ Desayuno, Almuerzo y Cena
    - Comercial
      - ◆ Sólo Desayuno
      - ◆ Desayuno, Almuerzo y Cena

*Modelo Facetado:*

- Entorno Natural (<--Faceta)
  - Ciudad (<--Categoría)
    - Ciudad hasta 500.000 habitantes (<--SubCategoría)
    - Ciudad superior a 500.000 habitantes (<--SubCategoría)
  - Montaña (<--Categoría)
  - Playa (<--Categoría)
- Perfil del Cliente (<--Faceta)
  - Vacacional (<--Categoría)
  - Convenciones y Eventos (<--Categoría)
  - Comercial (<--Categoría)
- Servicio de Comida (<--Faceta)
  - Desayuno (<--Categoría)
  - Desayuno, Almuerzo y Cena (<--Categoría)

Como se puede observar, en la versión facetada existen 3 facetas (Entorno Natural, Perfil del Cliente y Servicio de Comida), que a su vez contienen sus correspondientes categorías.

En este ejemplo se muestra la flexibilidad y facilidad para clasificar objetos mediante facetas en comparación con la clasificación jerárquica simple (aunque estas ventajas siempre dependen de la naturaleza de los objetos a clasificar y los objetivos de la clasificación).

Además, es posible observar que en el modelo facetado, en la faceta 'Entorno Natural', se ha añadido dos subcategorías a la categoría 'Ciudad', llamadas 'Ciudad hasta 500.000 habitantes' y 'Ciudad superior a 500.000 habitantes'. Incorporar estas categorías al modelo jerárquico supondría un enorme esfuerzo, ya que es un modelo menos flexible y escalable que el facetado.

A la hora de clasificar un hotel con este modelo habría que especificar a cuál o cuáles de las categorías de cada faceta corresponde o pertenece.