

Cátedra Tecnologías en comunicación Social.- Facultad de Periodismo y Comunicación Social - UNLP

**Lenguajes de marcado herramienta
dominante para la explicitación
formal y semántica**

Mela Bosch

melabosch@hotmail.com

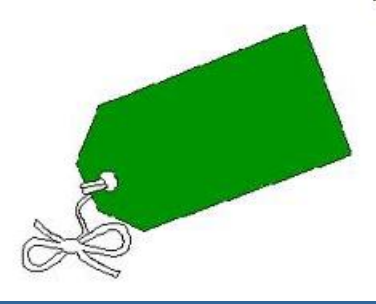


¿Qué son y para qué se usan las etiquetas?





*"Una etiqueta es una palabra clave o expresión pertinente asociada o asignada a una pieza de información (una imagen, un mapa geográfico, una entrada de blog, un videoclip, etc.) que describe el tema y permite la clasificación y la búsqueda de información basada en palabras claves."
(Fuente:[Wikipedia](#))*



Origen usos y etimología

El término ETIQUETA apareció en Francia en el siglo XIV con la forma **estiquette** para designar una señal escrita fijada en una estaca.

Se relaciona con la palabra española 'estaca', del gótico *staka* (palo), y con *stick*, inglesa.

Otro origen posible del término ETIQUETA es el rótulo que se ponía en los paquete de causas judiciales, escrito en latín: *Est-hic-quaestio* ("Aquí está la cuestión" se corrompió a *est-hi-quaest*, ->*et-hic-quet*, y finalmente -> *étiquette*,



Etiquetas: punto de vista de la semiología

Siguiendo a Peirce, un índice es un signo determinado por su objeto dinámico en virtud de la relación real que mantiene con ese objeto. El uso de índices nos permite identificar: desde el índice de los libros a los menús de un restaurante o los rótulos o etiquetas de los envases en el supermercado y también permiten comparar y valorar Ej: los índices económicos.

Un ícono es un signo determinado por su objeto dinámico en virtud de su naturaleza interna. Un ícono es una dama o un caballero en la puerta de un baño público, si además el baño es para discapacitados se le agrega el ícono pertinente. La generación de íconos es un trabajo intencional y hoy en día está siendo estudiado y utilizado en ámbitos tan disímiles como la arquitectura, la publicidad o la propaganda política.

El símbolo por su parte no tiene ninguna relación con su objeto dinámico, está allí por una convención o costumbre o decisión social. Las banderas de los países son un símbolo, cuyo color puede tener algún sentido atribuido pero que esencialmente tiene ninguna relación con la nacionalidad de las personas. Los números y palabras son símbolos.



En síntesis una ETIQUETA es siempre METAINFORMACION

La metainformación puede tener diferentes objetivos que es importante diferenciar:

- 1. Contextualizar una información**
- 2. Localizar y-o aglutinar información con características similares**
- 3. Comentar, anotar, opinar sobre una información**

Cada uno de estos objetivos en el etiquetamiento requiere diferentes procesos cognitivos

1. Contextualizar una información: fue el primer uso de las etiquetas en informática: se llamaba tag (etiquetas) a las órdenes o comandos de tipo descriptivo escritas en corchetes angulares que se agregaban en los programas, y luego con este procedimiento se desarrolló el marcado de los documentos en SGML y HTML

Cuando etiquetamos para contextualizar identificamos aspectos formales: formatos, tipos: imágenes, textos, tipos de letras, acentos, idioma, etc.

Cognitivamente realizamos
un proceso ANALITICO



Procesos cognitivos en el etiquetado

2. Encontrar o aglutinar información con características similares:

Se basa en un parecido fenomenológico u objetivo entre aspectos o propiedades de la información

Cognitivamente se realiza una ABSTRACCION, típicamente va del todo a la parte, está dentro de la tradición de taxonomía y organización del conocimiento en la tradición científica positivista.

Se trata de una orientación

HEURISTICA





Procesos cognitivos en el etiquetado

3. Comentar, anotar, opinar sobre una información:

Los conceptos que se relacionan se basan en aspectos **VALORATIVOS**

El proceso cognitivo está orientado a la interpretación, se trata de una orientación **hermenéutica** se usa para **recomendaciones y reputaciones**

**Riesgos: de la subjetividad a la intersubjetividad:
manipulaciones, sesgos**

○ **En el aspecto sociológico: estigmatización (Los autores de la teoría del etiquetamiento como estigmatización fueron, Becker, Lemert, Erikson, Kitsuse, Tannembaum. Schur y Gusfiel)**

¿Cuál es el uso las etiquetas?

A. Explicitación formal



Cuando publicamos una información en la web los motores de búsqueda utilizan los términos que **textualmente** aparecen, pero éstos pueden no ser interpretados en el mismo sentido por los sistemas, por eso es fundamental que cuando publicamos algo agreguemos palabras claves, es decir etiquetas que den cuenta de los conceptos fundamentales que queremos transmitir.

Al utilizar etiquetas hay que pensar como éstas pueden ser "*ganchos*" o "*rótulos*" eficaces para que tu contenido pueda ser más fácilmente buscado, encontrado y distribuido.

Las etiquetas representan aspectos de las páginas que son difíciles de capturar con términos de consulta normales en una búsqueda.

¿Cuáles son los pasos del etiquetamiento?

Se usa genéricamente el verbo etiquetar, o taguear, pero en realidad estamos ante dos procesos diferentes.

1. **Asignar, o bien es idear o pensar la etiqueta**, analizando si es útil y apropiada: **proceso conceptual**
2. **Insertar o asociar la etiqueta** en la pieza de información que se desea: **proceso físico**



¿Cuáles son los pasos del etiquetamiento?

¿Qué es idear una etiqueta?

Idear, o pensar la etiqueta, es la invención de una palabra o conjunto de palabras que mejor definan el contenido de lo que estamos publicando. Inventar etiquetas es una tarea delicada que exige analizar bien lo que se quiere transmitir. Para saber si una etiqueta ya existe debemos simplemente ponerla en los buscadores y ver qué resultados obtenemos.

¿Cómo idear o asignar una etiqueta?

Usar etiquetas requiere pensar en pasar de la búsqueda de información a la búsqueda de como se **nombra** una información. Para determinar cuál es la mejor etiqueta es necesario lograr conectarse con los usos de términos sobre el tema que nos interesa en el mundo, entonces es importante investigar en el tema y enunciar claramente el eje de la propuesta.



¿Como uso las etiquetas?

Explicitación formal



2. Debe estar relacionado con el contenido de la página: ¿cómo?

Criterio didáctico: el título debe contener palabras relevantes respecto al contenido de la página

- **palabras clave** que queremos **posicionar**, pero ojo, éstas deben estar contenidas también en los textos de tu página.
- Cuando alguien realice una búsqueda, este será el título que aparezca en los resultados. Si sobrepasamos los caracteres permitidos, el propio buscador lo cortará y el título aparecerá truncado, por lo que el orden de los términos en este caso sí tiene importancia.
- Si alguien realiza la búsqueda de una palabra que tenemos en nuestra etiqueta title y nuestra página aparece en los resultados, el title aparecerá como título en el resultado y la palabra clave en negrita, de ahí que sea muy importante incluir nuestras palabras clave en la etiqueta title, ya que aparecerá destacada.
- **Palabras claves: abstraer el contenido de lo general a lo particular.**

¿Como uso las etiquetas?

Criterio Semántico: Investigar los alcances de un término

- Polisemia
- Multidisciplina
- Multilingüísmo



¿Qué estrategias de asignación puedo usar?

Criterio Semántico: Lo mejor es usar etiquetas breves que no sean polisémicas. Cuando se enuncia la situación problemática se debe empezar a proponer las etiquetas y probarlas en la búsqueda y reformularlas de acuerdo con lo que aparezcan en los blogs.

Asociar las etiquetas a tu contenido de tal forma que se acerque lo más posible a las palabras que los usuarios potenciales usarían si hubieran encontrado tu contenido y se les pidiera que ellos lo etiquetaran.

¿Cómo empezar a asignar o idear etiquetas?



Criterio Semántico:

Empezar con lo básico: **identificar los conceptos** clave de una problemática, rastrearlos en la Internet, redes sociales, etc y reformularlos y luego etiquetar lo que se publica

Es decir llevar adelante un proceso bottom up, de prueba y error. La mejor manera de encarar la selección de etiquetas es analizar selectivamente los grandes grupos en los que puedes dividir tu información. Es lo que se denomina generar **categorías.**

¿Cómo saber si una etiqueta es verdaderamente útil y apropiada?

Ejercicio mental: pregúntase si a alguien que fuera a uno de los principales buscadores de sitios o de etiquetas como utilizará para su búsqueda de contenidos la etiqueta que estás considerando, y preguntate si le resultará útil el contenido que has asociado a dicha etiqueta.

Ejemplo: la etiqueta "países emergentes", cuando alguien utilice estas palabras para la búsqueda y encontrará mi nota, ¿este contenido será relevante? Cuanto más afirmativa sea la respuesta a esta pregunta, más probable será que la etiqueta que has seleccionado resulte adecuada.

Cuanto más genérica sea la etiqueta y menor el número de etiquetas asociadas a un determinado contenido, más difícil será definir exhaustivamente sus rasgos y características. Es por ello que la mayoría de los editores en línea suelen asignar tres, cuatro o más etiquetas a cualquier pieza de contenido que publican.

¿Cómo se insertan las etiquetas?

La mayoría de los sistemas de gestión de contenido tienen la forma de insertar etiquetas.

- Plataformas de blogs ya tienen integrada una herramienta para insertar etiquetas.
- Por ejemplo cuando se publica en un blog, al terminar el post se pueden poner las etiquetas.
- También hay plataformas que ya las tienen como parte fundamental ej YouTube

Se basan en los lenguajes de marcado



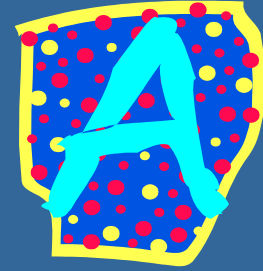
```
Document: Bungler OED      At: "<entry>"
<entry>
  <hwsec>
    <huwp>
      <hwlen>bungler</hwlen>
      <pron>b(I)></I>nglə</pron>. </huwp>
      <vfl>Also <vd>b</vd> <vf>bongler</vf>.
      </vfl>
      <etym>f. as prec. + <xra><xlen>-er</xlen>
    </hwsec>
    <res>One who bungles; a clumsy unskilful
    <quot>
      <qdat>1533 </qdat>
      <auth>None </auth>
      <uk>Assw. Poyson, 2k. </uk>Wks. (1557
      <qtxt>He is suen but a very bungler.
```

Lenguajes de marcado:
proceso físico:



Una gran solución y grandes problemas

El problema de la descripción y representación de las estructuras conceptuales en medios informáticos estuvo ligado y sobredeterminado por las necesidades de proceso → más importante lograr similitud con el lenguaje de programación que con el lenguaje de las personas



Precisiones terminológicas

Lenguaje: conjunto de símbolos normalizados para la representación y comunicación

Se dividen en: lenguajes naturales, lenguajes controlados y lenguajes artificiales.

Lenguas naturales: creadas en forma social y convencional conjuntos por individuos. Ej. lenguas de los comunidades humanas y formas de comunicación animal

Lenguajes controlados son porciones del lenguaje natural reglamentadas para fines específicos: lenguajes documentarios y las ontologías.

Lenguajes artificiales: conjuntos de símbolos arbitrarios también para fines específicos: los lenguajes de programación procesamiento, lenguajes de descripción, como el SGML, cuyo fin es la interpretación por parte de lenguajes de programación.



Lenguajes de marcado: fundamentos y principios

Extensiones

- **XML:** un dialecto de SGML contempla alteraciones estructurales respecto del lenguaje madre
- Es una versión extremadamente simplificada de efectos de crear un SGML genérico que permita procesar en Web objetos que no se pueden procesar con HTML: realidad virtual, movimiento, señales olfativas, etc.
- XML es un metalenguaje, ya que permite a su vez crear subconjuntos, de hecho se está reescribiendo todo HTML en XML.

Se normalizó la parte XML para libros, artículos y series con la norma ISO 12083 (última 2009)

Característica	1ª Generación	2ª Generación
Lenguaje principal	HTML	XML
Forma y estructura	documentos no estructurados	documentos estructurados
Semántica	semántica implícita	etiquetado explícito (metadatos, Web semántica)
Relación entre contenido y forma	HTML = fusión de forma y contenido	estructura en capas de forma y contenido: XML + transformación (p.e., XSL) a HTML, WML, PDF, u otros formatos
Editabilidad	documentos estáticos	documentos dinámicos
Descomponibilidad y recomponibilidad	sitios Web monolíticos, independientes	bricolaje (agregación), sindicación, reasignación de contenido
Interactividad	medio de difusión unidireccional	Web editable, bidireccional
Audiencias	para consumo humano	para humanos y ordenadores (p.e., servicios Web)
Control de producción	centralizado	descentralizado (peer-to-peer P2P)

Primera web:

Navegación en páginas y el acceso a datos estructurados través de bases de datos en línea

La Web 2.0:

Arquitecturas interactivas, interoperables y de alta usabilidad

Ambas posturas persisten: el control y el espontaneísmo

La Web 3.0:

Incluye las precedentes
territorio diferente de las épocas previas
más allá del optimismo inicial
enorme esfuerzo:

- ✓ Conceptualizaciones
- ✓ Trabajo colectivo para sinergia en los modos de organización y circulación del conocimiento





Qué es lo que hay hoy en la Web

- **Web superficial:** navegable y accesible desde sus páginas
 - **Web profunda:** datos estructurados controlados por sistemas gestores de bases de datos
- La mayor parte de los sistemas que acumulan información de manera persistente, incluidos muchos de los de referencias bibliográficas y bibliotecas digitales se basan en datos estructurados con un modelo dominante que es el Bases de datos relacionales.

Google

Yahoo

msn

Surface
visible web

(~20 %)

Deep
invisible web

(~80 %)

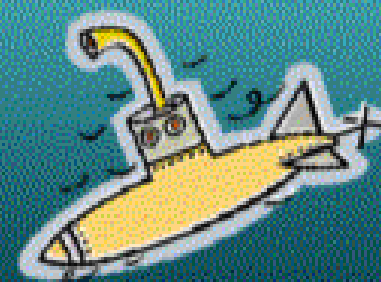
Specialized databases with
dedicated search engines

Non-conventional formats

Paid information

Dynamic webpages

Restricted access information



Lenguajes de marcado: como se usan



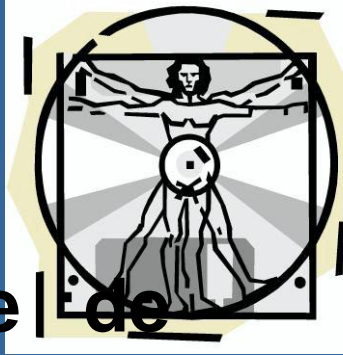
Aspectos de software

Es posible codificar textos en SGML en cualquier editor de texto, la tarea artesanal es muy lenta y pasible de errores

Tres tipos de software

- Editores
- analizadores (parsers)
- visualizadores (browsers)

El futuro: la Web Semántica



Web Semántica: usa el lenguaje de marcado tanto para expresar los datos como para darles sentido aunque tengan diferente origen y forma

Tres tecnologías Web Semántica que se valen del lenguaje de marcado:

- XML
- Resource Description Framework (RDF)
- Ontologías.

¿Qué es Web Semántica?

Síntesis:

Una sintaxis de datos sobre datos (metadatos) para dar semántica

Para que los servidores y navegadores pueden mostrar los datos al usuario basados en su contenido y no solo en su estructura

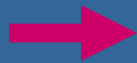


Metodología

Etiquetado que describe el contenido que encapsula



**Semantic
Web**



indiza recursos web: objetos,
no materiales, sino objetos digitales

Esos Objetos son descritos a través de metadatos que están especificados en

- una ontología (enfoque top down)
- Esquemas de metadatos (enfoque bottom up)

Los metadatos requieren un fundamento lógico que tiene dos aspectos:

- ✓ Una Lógica descriptiva: instrumentada por lenguajes de marcado
- ✓ una Lógica de objetos

Una sola herramienta: la anotación semántica: etiquetamiento altamente preciso

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN" "http://www.w3.org/TR/html4/loose.dtd">
<html>
<head>
<meta name="TITLE" content="...">
<meta name="KEYWORDS" content="...">
<meta name="DESCRIPTION" content="...">
<link rel="stylesheet" type="text/css" href="...">
<script language="javascript" type="text/javascript">
</head>
<body bgcolor="#ffffff" width="100%">
```