

# TRABAJO ROBOTS

---

## DESARROLLO WEB EN ENTORNO CLIENTE

Pablo Cidon Barrio

Curso 2017-218

## Contenido

Definición .....	3
Clasificación.....	3
Tipos de robots de búsqueda.....	3
Funcionamiento .....	3
Etiqueta <meta>.....	4
Atributos.....	4
Cronología .....	5
Referencias Web .....	5

## Definición

Es un robot que recorre la estructura del hipertexto buscando los enlaces que se encuentran en el mismo, para guardarlos en las bases de datos de los motores de búsqueda. Al contrario, los Directorios suelen formarse de forma manual por los usuarios y no recuperan automáticamente los enlaces que se encuentran en las páginas web, sino que se limitan a encontrar lo que las personas incluyen en ellos, pudiendo clasificar por secciones los temas de las páginas web.

## Clasificación

Principales buscadores o motores de búsqueda de Internet.

Se pueden clasificar en tres tipos:

- **Buscadores jerárquicos:** interfaces de interrogación textual contra bases de datos de representaciones de páginas web, creadas según un modelo vectorial.
- **Directorios:** directorios de enlaces a páginas (agrupando sus enlaces por categorías) que ofrecen motores de búsqueda interna.
- **Metabuscadore:** interfaces de reenvío de búsqueda a múltiples buscadores.

## Tipos de robots de búsqueda

- **Arañas:** se usan para rastrear la red. Lee el hipertexto y accede a todos los enlaces a los que se hacen referencia desde el mismo.
- **Gusanos:** es lo mismo que una araña, pero es una réplica del programa a diferencia de la araña, que es un programa original. Se usan para duplicar los directorios FTP para poder acceder a más usuarios.
- **Orugas:** es un tipo de robot que ha dado lugar el nombre de algunos buscadores como por ejemplo [WebCrawler](#).
- **Hormigas:** consiste en una cooperativa de robots. Trabajan distribuidamente, explorando diferentes partes de la web. Siguen un mismo objetivo.
- **Vagabundos:** es un tipo de robot que realiza estadística en la web.
- **Robots de Conocimiento:** localizan referencias de hipertexto dirigidas a un documento o a un servidor. Permiten la evaluación del impacto de las distintas aportaciones que forman las distintas áreas de conocimiento de la web.

## Funcionamiento

En la mayoría de los buscadores se usan las arañas. Estas recorren las páginas extrayendo información sobre los contenidos de las páginas. Cuando se hace una búsqueda en los motores, estos consultan en su base de datos y devuelven los datos clasificados según su importancia. De las webs, los buscadores guardan desde la página de entrada a todas las que se encuentran en el servidor.

Si se busca por una palabra concreta, ofrecerá como resultados páginas que contengan esta palabra en alguna parte del texto. Si consideran que la página puede ser importante para el usuario, almacenan todas las páginas, pero si no, solo almacenan aquellas páginas que pueden ser importantes.

Cada cierto tiempo, los motores revisan los sitios para actualizar su base de datos, por lo que en ocasiones los contenidos no estén actualizados.

Los robots son utilizados para varios propósitos:

- Indexar.
- Validar HTML.
- Validar links.
- Monitorizar.
- Generar imágenes.

### Etiqueta <meta>

Esta etiqueta sirve para aportar información acerca del documento.

Solo requiere de una etiqueta "<meta (atributos)/>" que finaliza con autocierre.

Es un elemento de cabecera, que únicamente se puede colocar en la etiqueta <head>.

Las meta *tags* son unas etiquetas pertenecientes al HTML que se deben escribir dentro del *tag* general <head> y que lo podemos definir como líneas de código que indican a los buscadores que indexan la página por qué términos debe ser encontrada. Dependiendo de la utilización, caracterización y objetividad de dichos *meta*, más otros factores seo se puede conseguir una excelente posición en el listado resultante de una búsqueda.

### Atributos

Atributo	Descripción	Valor
Genéricos		
lang	Información sobre el idioma del contenido y el valor de sus atributos.	Un código de idioma, fijado por el navegador.
dir	Indica la dirección del texto y tablas.	'ltr' ó 'rtl' es fijado por el navegador.
Específicos		
name	Nombre que se asocia a la información.	Distingue mayúsculas y minúsculas, fijado por el navegador.
content	Datos que se quieren asociar a la etiqueta name.	Texto, también distingue mayúsculas y minúsculas.
http-equiv	Aporta información sobre el encabezado de respuesta HTTP.	Por defecto, lo fija el navegador.
scheme	Indica un esquema de interpretación de metadatos.	Por defecto lo fija el navegador. También distingue entre mayúsculas y minúsculas.

## Cronología



## Referencias Web

[Clasificación](#)

[Funcionamiento y tipos de robots](#)

[Etiqueta <meta>](#)

[Cronología](#)