

# RECUPERACIÓN DE LA INFORMACIÓN EN INTERNET

MSc. Patricia Raquel Castillo Aranibar

# RECUPERACIÓN DE LA INFORMACIÓN EN INTERNET

## 1. Buscadores Web

- Tipología
- Componentes y funcionamiento
- Ejemplos de Motores de Búsqueda
- Buscadores Web especiales

## 2. Internet Invisible

## 3. Posicionamiento en Recuperación Web

## 4. Optimización de Páginas Web

## 5. Criterios de Evaluación de buscadores

# ¿QUÉ ES UN BUSCADOR WEB?

Un **buscador** (*search engine*) es un software que recopila e indexa archivos almacenados en servidores web y recupera conforme a algunos criterios específicos.



## Objetivos

- Recopilar la información de la red
- Indizar la red constantemente para permitir la consulta de sus índices
- Encontrar los documentos que contengan las palabras clave introducidas por el usuario.

## Tipos:

- Directorios o índices temáticos
- Motores de búsqueda
- Metabuscadores

# TIPOLOGÍA: DIRECTORIOS

## ¿Qué es un Directorio?

- Sitio Web que gestiona una BBDD gestionada según criterios manuales
- Las URL están clasificadas en categorías temáticas
- Suele estar organizado por temas y categorías jerárquicas

## Características

- Selección y clasificación manual de recursos
- Datos poco actualizados y poco exhaustivos
- Resultados relevantes y páginas de calidad
- Suelen ser temáticos

# TIPOLOGÍA: DIRECTORIOS

- ❑ Argus Clearinghouse
- ❑ DMOZ (The Open Directory Project -ODP- o Proyecto de Directorio Abierto)
- ❑ Yahoo!
- ❑ Google Directory
- ❑ Buscador.com <http://www.buscador.com>
- ❑ Lasonet.com [www.lasonet.com](http://www.lasonet.com) es un directorio temático de enlaces a las páginas de Internet organizadas por temas, prensa, bancos, pueblos, buscadores, planos, etc.
- ❑ Web Directorio <http://www.web-directorio.com>
- ❑ Bubl <http://bubl.ac.uk/> (Con la clasificación Dewey)
- ❑ Galaxy <http://www.galaxy.com>
- ❑ IPL the Internet Public Library <http://ipl.org/ref>
- ❑ Intute (RDN) <http://www.intute.ac.uk>

# DIRECTORIOS TEMÁTICOS: YAHOO!

- Catorce materias subdivididas en un número similar de subtemas. Bueno para Usabilidad.
- Se puede hacer una búsqueda general en cualquier sección o nivel. Si no encuentra resultados “salta” Yahoo!Search
- Cada resultado consiste en un título o una breve descripción.

Directorio de sitios Web ¡Hemos organizado la web para ti!

## Arte y cultura

[Literatura](#), [Teatro](#), [Museos](#), [Moda](#)...

## Internet y ordenadores

[WWW](#), [Software](#), [Chat](#), [Redes](#)...

## Ciencia y tecnología

[Astronomía](#), [Biología](#), [Ingeniería](#)...

## Materiales de consulta

[Bibliotecas](#), [Diccionarios](#)...

## Ciencia Sociales

[Filosofía](#), [Historia](#), [Idiomas](#), [Psicología](#)...

## Medios de comunicación

[Radio](#), [TV](#), [Revistas](#), [Periódicos](#)...

## Deportes y ocio

[Coches](#), [Fútbol](#), [Videojuegos](#), [Turismo](#)...

## Política y gobierno

[Elecciones](#), [Boletines oficiales](#), [Ministerios](#)...

## Economía y negocios

[Empresas](#), [Inmobiliarias](#), [Empleo](#)...

## Salud

[Medicina](#), [Enfermedades](#), [Embarazo](#)...

## Educación y formación

[Primaria](#), [Secundaria](#), [Universidades](#)...

## Sociedad

[Gastronomía](#), [Religión](#), [Para niños](#)...

## Espectáculos y diversión

[Actores](#), [Música](#), [Humor](#), [¡Genial!](#)...

## Zonas geográficas

[Países](#), [Europa](#), [España](#), [CC.AA.](#)...

[Los más buscados de 2003](#) - [Novedades](#)

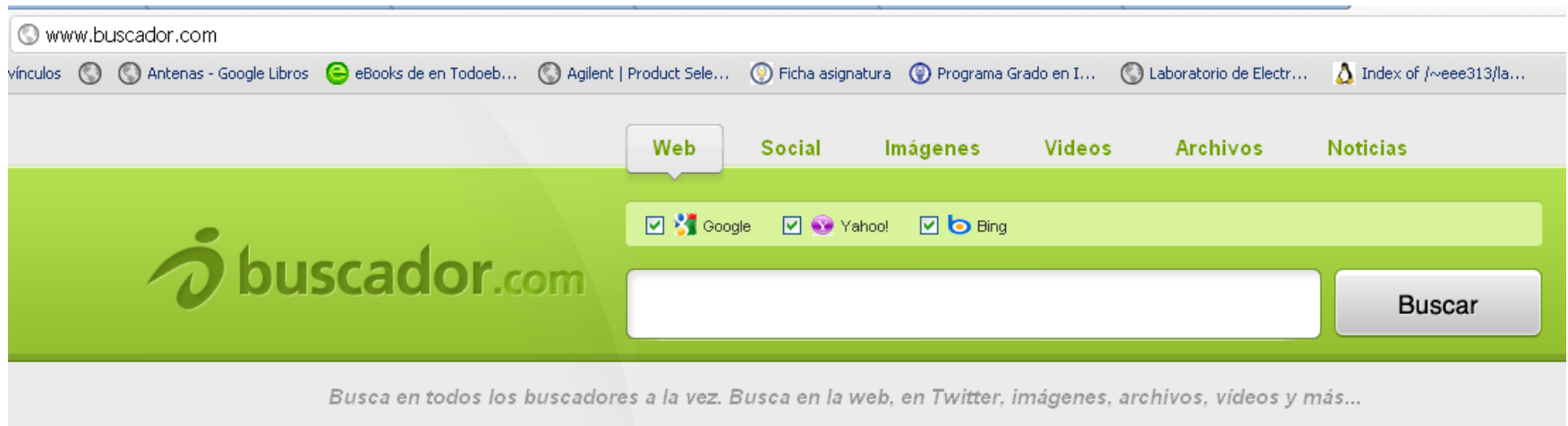
# DIRECTORIOS TEMÁTICOS: DMOZ/ OPEN DIRECTORY PROJECT



The screenshot shows the DMOZ Open Directory Project homepage. At the top, there is a green header with the DMOZ logo and the text "open directory project". To the right, it says "in partnership with AOL" and "search". Below the header, there is a navigation bar with links: "qué es dmoz", "bitácora de dmoz", "sugerir URL", "reportar abuso/spam", and "ayuda". A search bar is located below the navigation bar, with a "Buscar" button and a dropdown menu set to "en todo el directorio". The main content area is titled "Top: World: Español (168.312)" and has a "Descripción" link. Below this, there is a list of categories and their counts:

- [Regional](#) (109.439)
- [Artes](#) (17.561)
- [Ciencia y tecnología](#) (4.330)
- [Compras](#) (894)
- [Computadoras](#) (3.518)
- [Deportes](#) (1.308)
- [Educación](#) (4.977)
- [Hogar](#) (524)
- [Juegos](#) (1.808)
- [Medios de comunicación](#) (1.207)
- [Negocios](#) (3.154)
- [Referencia](#) (1.362)
- [Salud](#) (2.561)
- [Sociedad](#) (10.734)
- [Tiempo libre](#) (4.935)
- [Niños y jóvenes@](#) (2.439)
- Usenet esp.bienvenida - [news](#) - [Google Groups](#)
- Buscar "Español" en: [AltaVista](#) - [A9](#) - [AOL](#) - [Ask](#) - [Chusty](#) - [Gigablast](#) - [Google](#) - [Lycos](#) - [MSN](#) - [Yahoo](#)

# DIRECTORIOS TEMÁTICOS: Características de Buscador



## Lo más buscado

4 amigos amor angel barcelona casa facebook febrero goglee google googlee imagenes  
madrid pareja por real madrid tipos de interes trabajos ver peliculas online www.gogle.com

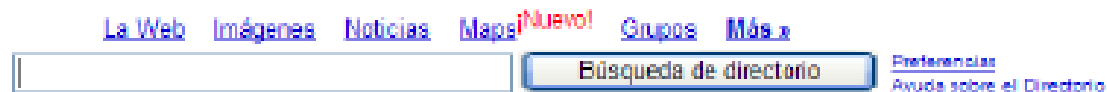
## Lo más reciente

agregar barra de estado agregar barra de estado carga agregar barra de estado carga acces 2007 agregar barra de estado  
carga acces 207 agregar barra de progreso carga acces 2007 decoracion de salones pintura harenatv.com importar Vinculada  
de ODBC whale-tail.com iE%ē™¼ē±ē|ē², 550612

Directorio | Agréganos a tus buscadores | Antevenio 2011



# DIRECTORIOS TEMÁTICOS: Características de Google



La Web organizada por temas en categorías.

## Artes

[Cine](#), [Música](#), [Diseño](#), [Fotografía](#), ...

## Negocios

[Construcción](#), [Empleo](#), [Turismo](#), ...

## Computadoras

[Hacking](#), [Multimedia](#), [Redes](#), [Usenet](#), ...

## Juegos

[Apuestas](#), [Internet](#), [Videojuegos](#), ...

## Salud

[Enfermedades](#), [Farmacia](#), ...

## Hogar

[Bricolaje](#), [Cocina](#), [Familia](#), ...

## Medios de comunicación

[Digitales](#), [Radios](#), [Televisión](#), ...

## Tiempo libre

[Autos](#), [Mascotas](#), [Motos](#), [Nudismo](#), ...

## Referencia

[Diccionarios](#), [Mapas](#), [Refranes y dichos](#), ...

## Regional

[África](#), [América](#), [Asia](#), [Europa](#), ...

## Ciencia y tecnología

[Biología](#), [Ciencias Sociales](#), [Química](#), ...

## Compras

[Arte](#), [Informática](#), [Regalos](#), [Vestuario](#), ...

## Sociedad

[Filosofía](#), [Gente](#), [Religión](#), [Sexualidad](#), ...

## Deportes

[Atletismo](#), [Fútbol](#), [Golf](#), ...

Categoría relacionada:

[Arte](#) [Cine](#) [Cocina](#) [Cultura](#) [Educación](#) [Empleo](#) [Entretenimiento](#) [Familia](#) [Fotografía](#) [Gente](#) [Hogar](#) [Informática](#) [Internet](#) [Música](#) [Negocios](#) [Religión](#) [Salud](#) [Sociedad](#) [Televisión](#) [Turismo](#) [Vestuario](#)

# TIPOLOGÍA: Motores de búsqueda

## ¿Características de los motores de búsqueda?

- Software con un sistema de recolección de URLs e indexación automatizadas
- Generales o temáticos (buscadores verticales)

## Características

- Muy exhaustivos
- Muy actualizados
- Manipulables
- Problemas con la calidad de los resultados y ambigüedad semántica

# TIPOLOGÍA: Motores de búsqueda

## Ejemplos

- Google
- MSN - Microsoft
- Teoma/AskJ
- Yahoo! Search
- Exalead
- Virgilio.it
  
- Searchenginewatch <http://www.searchenginewatch.com/>

# TIPOLOGÍA: Motores de búsqueda

- ❑ **Generales** (Google, Alexa, Ask, Exalead, Wikia Search, Yahoo!, Msn, Yandex (ruso <http://www.yandex.ru>) Baidu (Chino <http://www.baidu.com>) Cuil (<http://www.cuil.com>))
- ❑ **Temáticos**
  - ❑ De alcance limitado geográficamente
  - ❑ Financieros
  - ❑ De negocios
  - ❑ Para la empresa
  - ❑ Dispositivos de búsqueda
  - ❑ Empleo
  - ❑ Legal
  - ❑ Médicos
  - ❑ Noticias
  - ❑ Personas
  - Propiedad inmobiliaria

# TIPOLOGÍA: Motores de búsqueda

- ❑ Temáticos
  - ❑ Video juegos
  - ❑ Foros
  - ❑ Blogs
  - ❑ Multimedia
  - ❑ Código fuente
  - ❑ BitTorrent
  - ❑ Email
  - ❑ Mapas
  - ❑ Precio
  - ❑ Preguntas y respuestas (answers.com)
  - ❑ Motores de búsqueda de código abierto (Google Code Search, Jexamples, Koders, Krugle)
  - ❑ Motores de búsqueda sociales
  - ❑ Motores de búsqueda visuales
  - ❑ Usenet
- Buscadores basados en otros buscadores

# TIPOLOGÍA: Motores de búsqueda

## Extintos

- ❑ BRS/ Search
- ❑ Google Answers
- ❑ Infoseek
- ❑ Inktomi
- ❑ Lotus Magellan
- ❑ Noxtrum
- ❑ Overture.com
- ❑ PubSub
- ❑ Singingfish
- ❑ WiseNut
- ❑ World Wide Web Worm

## Transformaciones

- ❑ Web Crawler (ahora metabuscador)
- ❑ Excite (ahora metabuscador)
- ❑ Teoma /AskJ

# TIPOLOGÍA: Metabuscadores

## ¿Qué son?

- Software que agrega los resultados de varios motores o directorios para encontrar las páginas más relevantes.

## Características

- Sin base de datos propia
- Optimización por tiempos de respuesta
- Incertidumbres sobre métodos de combinación de buscadores, pesos, orden de resultados, ...

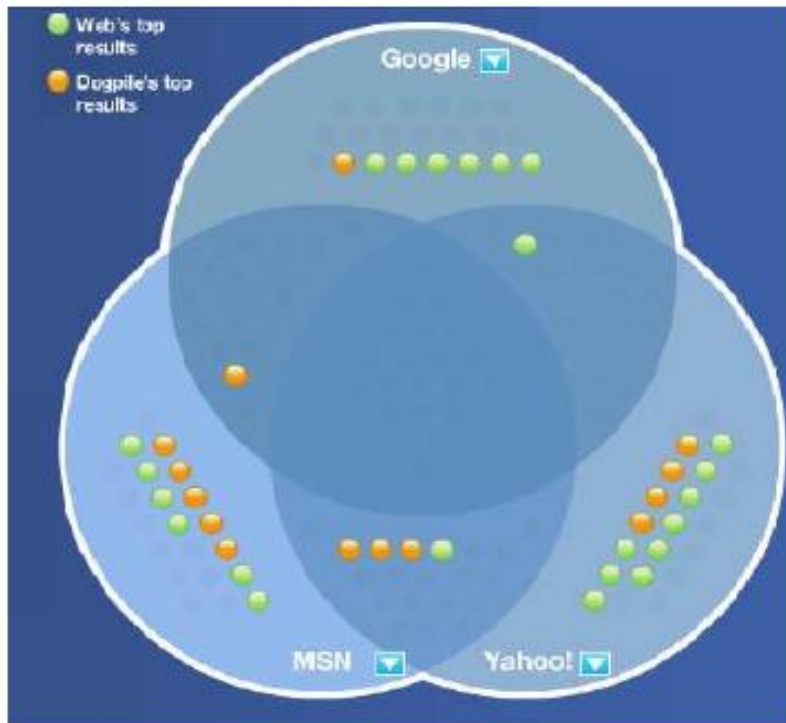
## Tipos:

- Metabuscadores propiamente dichos
- Multibuscadores
- Agentes de búsqueda

## Ejemplos

- MetaCrawler, Mibúsqueda, Copernic

# TIPOLOGÍA: Metabuscadores



**Ejemplo:** comprueba cuantas páginas únicas existen entre las primeras de cada buscador [Google, Yahoo, MSN y AJ] en promedio es un 1%

Comparar





# TIPOLOGÍA: Metabuscadores

- ❑ Metabuscadores: combinan los resultados, lanzando una consulta en varios buscadores. Ejemplo: MetaCrawler
- ❑ Multibuscadores: no combinan los resultados, sólo lanzan la consulta en varios buscadores. Ejemplo: My Search ([MiBúsqueda](#))
- ❑ Agentes de Búsqueda: metabuscadores instalado localmente. Ejemplo: WebFerrer, Copernic

# TIPOLOGÍA: Metabuscadores

- ❑ **MetaCrawler** [www.metacrawler.com](http://www.metacrawler.com) *Elimina los duplicados*
- ❑ **Dogpile** [www.dogpile.com](http://www.dogpile.com) *motores distintos según categoría*
- ❑ **Vivisimo** [www.vivisimo.com/](http://www.vivisimo.com/) *con clusters y posición en cada buscador*
- ❑ **Kartoo** <http://www.kartoo.com/> *mapas de **clusters** navegables y con **expansión de consultas***
- ❑ **Myway** <http://myway.com/>
- ❑ **SurfWax** <http://www.surfwax.com/> *(en la opción focus con **expansión de consultas en inglés mediante tesauro**)*
- ❑ **Ixquick** <http://www.ixquick.com/> *buen metabuscador (cada estrella un buscador) con **refinamiento** de búsqueda*
- ❑ **Beaucoup** <http://www.beaucoup.com/> *combina un metabuscador con un directorio y con términos específicos)*

# RECUPERACIÓN DE LA INFORMACIÓN EN INTERNET

## 1. Buscadores Web

- Tipología
- Componentes y funcionamiento
- Ejemplos de Motores de Búsqueda
- Buscadores Web especiales

## 2. Internet Invisible

## 3. Posicionamiento en Recuperación Web

## 4. Optimización de Páginas Web

## 5. Criterios de Evaluación de buscadores

# FUNCIONAMIENTO MB: Componentes



## Componentes

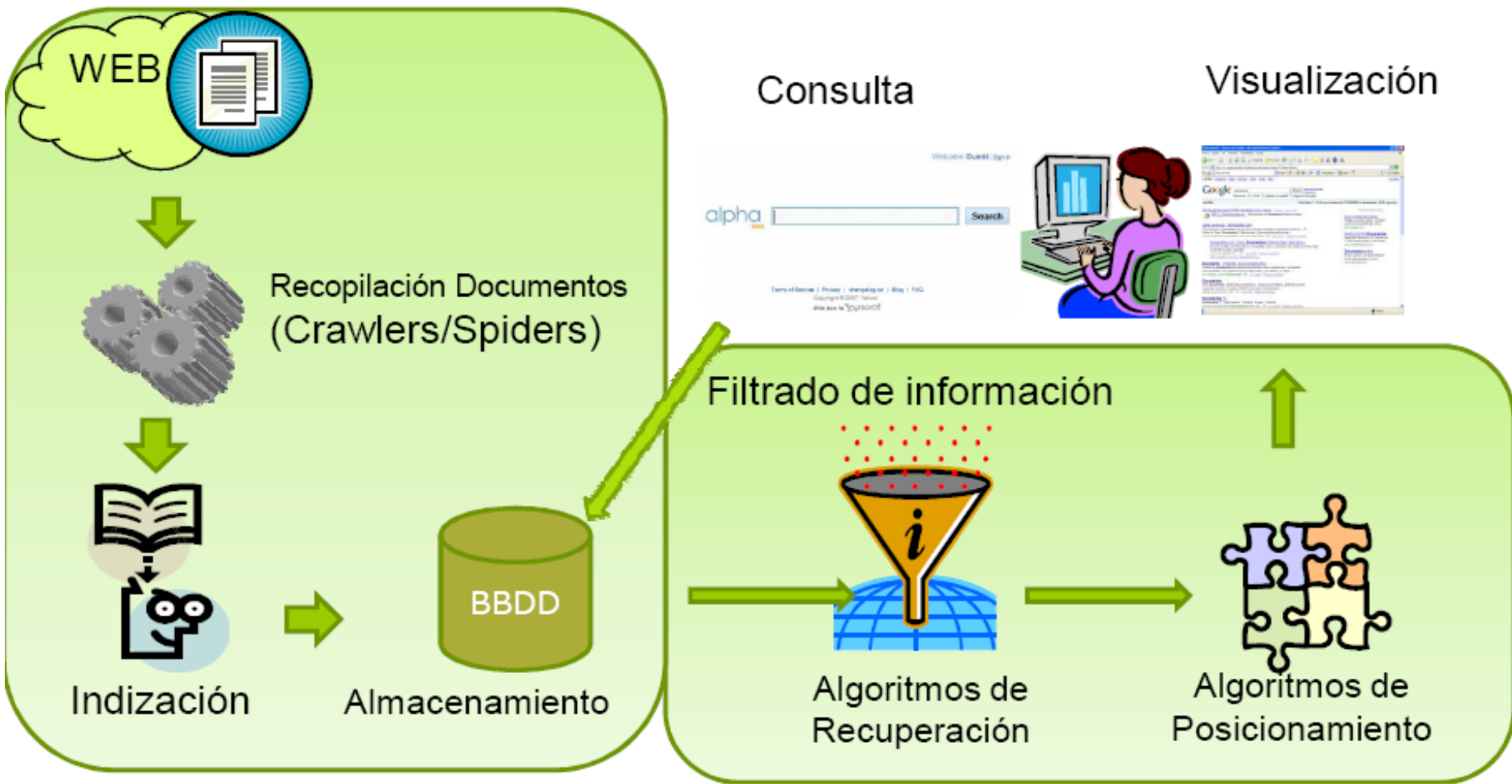
- Spider/Robot/Crawler. Robot.txt para ver directorios permitidos: Localizador y Recolector.
- Indizador
- Base de datos
- Interfaz de búsqueda
- Interfaz de Recuperación



## Cada Buscador tiene su propio motor:

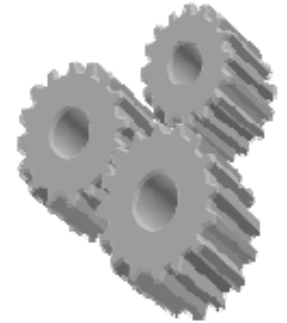
- ✓ Altavista-Scooter,
- ✓ Lycos-Tres,
- ✓ Excite-Architext,
- ✓ Infoseek-Sidewinder,
- ✓ Google-Googlebot,...

# Motores de Búsqueda WEB: Funcionamiento



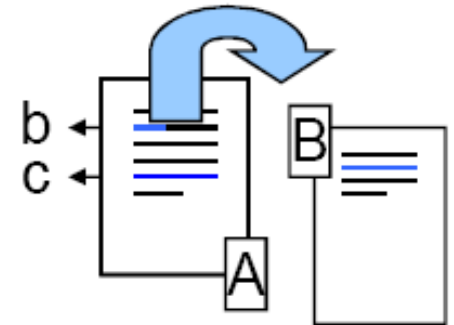
# Motores de Búsqueda: Recopilación de documentos

*Spider, crawlers, robots* o Agentes de Búsqueda son los nombres que recibe el software que recopila los documentos.



## Funcionamiento

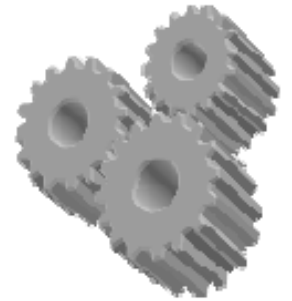
- Comienza en una página (A) y recopila todas sus URL
- Envía la página (A), comprueba que no está indizada y que no se tiene una versión menos actualizada, indiza la página (A)
- Recupera la página (B) que está primera en la lista
- Envía la página (B)...



# Motores de Búsqueda: Recopilación de documentos

## Criterios para organizar la lista a procesar:

- Puede tener en cuenta novedad o prestigio.
- Depth First Crawling: Hasta que no acaba con todas las páginas de un site no pasa a las del siguiente site.
- Breath Crawling: procesa primero las primeras páginas que ha encontrado en cada site, luego las segundas páginas de cada site, etc.



# Motores de Búsqueda: Base de datos



## ■ Grandes bases de datos:

- Google
- Yahoo
- MSN
- Teoma
- Wisenut/LookSmart
- Gigablast
- Exalead / Quaero

Los demás buscadores utilizan estas BD.



# Base de datos: Ficheros inversos

- Permite búsquedas rápidas de textos
- Cada término asociado a un conjunto de URLs y opcionalmente a su frecuencia y posición en cada URL.
  - Suele dividirse en glosario, con la frecuencia total y número de documentos
  - Y resto de datos: URL, posición o frecuencia en el documento.
- El glosario puede estar repetido y repartido en distintas máquinas.
- Una opción para acelerar es: las páginas más populares pueden estar en muchos servidores (y a ellos se acude primero), si no hay resultados se acude a unos pocos servidores que tienen las menos populares



Término	#d	#frc
A	99	128
Arandela	1	2
Baraja	2	2
Casa	31	40
Tecla	1	1
...	..	..
...	..	..

LEXICON

URL	Posición
uc3m.es	34, 45, 78
uc3m.es	43, 67
cartas.com	45
<a href="http://www.o.org">www.o.org</a>	33
...	..
...	..
...	..
...	..

POSICIÓN (POSTINGS)

# Base de datos: Ficheros inversos

Término	#d	#frc
A	99	128
Arandela	1	2
Baraja	2	2
Casa	31	40
Tecla	1	1
...	..	..
...	..	..

LEXICON

URL	Posición
uc3m.es	34, 45, 78
uc3m.es	43, 67
cartas.com	45
<a href="http://www.o.org">www.o.org</a>	33
...	..
...	..
...	..
...	..

POSICIÓN (POSTINGS)



# Ejemplo: La base de GOOGLE

## Característica de la base de datos de Google

- “+” antes de una palabra no elimina aun siendo vacía, si se quiere buscar por frase poner comillas. “-” que no aparezca un término.
- No es lo mismo la ubicación geográfica desde donde hagamos la consulta (desde 2004)
- El orden de las palabras importa
- La misma consulta desde un mismo sitio con intervalo de segundos puede dar resultados distintos.
- No admite truncamiento, es decir poner singular y plural
- No distingue mayúsculas, poner sin acentos
- Búsquedas por campos limitado
- Imposible combinar operadores booleanos de carácter distinto (todos AND todos OR pero no paréntesis)
- Aunque Google diga que existan 2000 resultados, jamás podrás pasar del resultado 1000.



# Motores de Búsqueda: opciones búsqueda por campos en Google

## Opciones de búsqueda por campos

- Descubrir vínculos que le apuntan link:www.google.com
- restricciones de búsqueda de un dominio "site:ejemplodedominio.com", "site:information"
- para encontrar información de artículos de prensa en el sitio de Google: press site:www.google.com
- Para que aparezca en el título: intitle, allintitle
- Para que aparezca en la url: inurl, allinurl
- Definition: "palabra"



## "Betas de siempre" de Google

- MoreGoogle, GoogleDesktop, Barra de Google
- Google Scholar, APIS (recuerda que van contra un servidor no actual)
- Calculadora
- Google Suggest, profiles y demás google labs
- Y demás Gmail (los indiza para adwords), telefonía IP, ..

# RECUPERACIÓN DE INFORMACIÓN EN INTERNET

## 1. Buscadores Web

- Tipología
- Componentes y funcionamiento
- Ejemplos de Motores de Búsqueda
- Buscadores Web especiales

## 2. Internet Invisible

## 3. Posicionamiento en Recuperación Web

## 4. Optimización de Páginas Web

## 5. Criterios de Evaluación de buscadores

# Motores de Búsqueda: Yahoo Search

→ Yahoo! fue el primer gran directorio hecho a mano, tardo cinco años en llegar al millón de páginas, mucha calidad, poca actualización, poco exhaustivo.

2004: compra AltaVista (por contenido de su BD), AlltheWeb (por contenido de su BD), Inktomi (por su algoritmo de posicionamiento y estructura de BD), Kelkoo (como comparador de precios) y Oberture

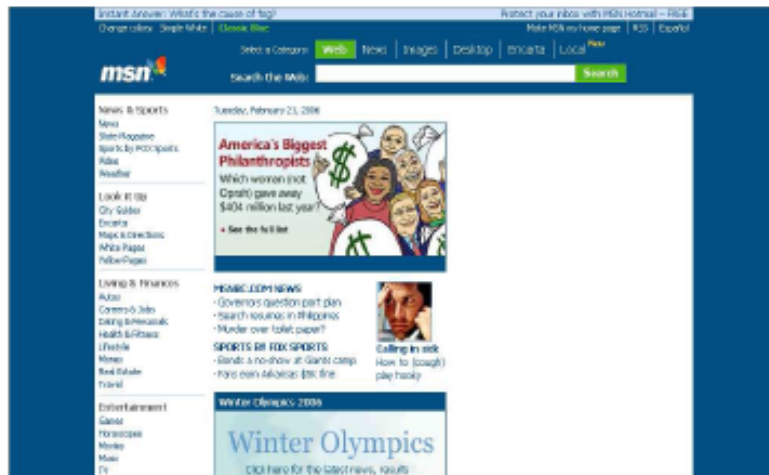
→ Comienza a utilizar Yahoo Search, con el motor de Inktomi

En 2004 muchos problemas con integración de BD, se quejan mucho de la calidad resultados...mejora mucho

Posicionamiento: asignación de pesos denominada WebRank (parece relacionada con la barra de búsqueda), el interfaz y otros pesos de posicionamiento copiados de Google

# Motores de Búsqueda: MSN

- Copia el algoritmo de posicionamiento de Google en 2004
- Problemas por instalación por defecto en las aplicaciones MS
- Actualmente el mejor valorado junto con Google y Yahoo!



# Ejemplos: A9 y AskJeeves

## ■ A9

- Permite buscar a texto completo en libros de muchas editoriales
- Pertenece a Amazon

## ■ AskJeeves

- Ha comprado a Excite, iWon y a Teoma
- Siempre ha tenido algo de PLN y ha establecido comparación con news
- Actualmente el motor de Teoma hace que AJ tenga los mejores rankings de precision. Teoma utiliza un criterio denominado autoridad basado en los enlaces de las páginas del mismo tema que apuntan a la página.
- Es actualmente el cuarto buscador mundial



# RECUPERACIÓN DE INFORMACIÓN EN INTERNET

## 1. Buscadores Web

- Tipología
- Componentes y funcionamiento
- Ejemplos de Motores de Búsqueda
- Buscadores Web especiales

## 2. Internet Invisible

## 3. Posicionamiento en Recuperación Web

## 4. Optimización de Páginas Web

## 5. Criterios de Evaluación de buscadores

# Buscadores Web especiales

1. Por su temática
2. Por su tecnología y recursos
3. Por los servicios que ofrece

# Búsqueda de Código

## ❑ Koders <http://www.koders.com/>

The screenshot shows the Koders website interface. At the top, there's a navigation bar with links: Company, Plugins, Support, Community, and My Account. Below this is the Koders logo, which includes the text 'koders' and 'powered by blacklucksoftware'. A status line indicates 'Searching 1,531,726,841 lines of open source code.' The main search area features a large input field containing 'nlp' and a 'Search' button. Below the input field are two dropdown menus: 'All Languages' and 'All Licenses', along with an information icon. A summary bar shows '26 Interface Definitions, 176 Class Definitions, 5 Method Definitions'. The search results section displays 'Results 1-25 of about 1,848 results found for 'nlp' in 0.66 seconds'. A pagination bar shows '1 2 3 4 5 6 7 8 9 10 ...'. The first result is 'LexicalMWE.java', showing its source code snippet: 

```
// Foundation, Inc., 59 Temple Place - Suite 330, Boston, MA 02111-1307, USA.
////////////////////////////////////
package opennlp.grok.preprocess.mwe;
import java.io.*;
import java.util.*;
```

 Below the code, it lists 'Language: Java', 'License: LGPL', '(C) 2002 Mike Atkinson', 'LOC: 61', and the 'SourceForge' link. The second result is 'libVers.pl', showing its source code snippet: 

```
package Vers;
* for logfiles, etc.
```

 On the right side, there's a 'Project Matches' section with a list of projects: 'Text\_NLP - Text::NLP', 'nlp', 'Infomap NLP Software', 'java-nlp - Natural Language Processor', and 'HLI Medical NLP Toolkit'. At the bottom of this section, it says 'Can't find a project? Add it to our index'.

# Búsqueda de Código

- Google Code Search (<http://www.google.com/codesearch>)

 [Búsqueda avanzada de código](#)

## Búsqueda de código fuente público

Sintaxis y ejemplos ([más información sobre expresiones regulares](#))

<code>regexp</code>	Buscar una expresión regular <code>go(2)gle bello, world, int printf</code>
<code>"exact string"</code>	Buscar una cadena exacta <code>"compiler happy"</code>
<code>class:regexp function:regexp</code>	Buscar únicamente en nombres de clases y funciones <small>Nuevo!</small> <code>class BTree function laugh</code> <code>class hash "multimap function: range"</code>
<code>file:regexp</code>	Buscar sólo en los archivos o directorios que coincidan con <code>regexp</code> <code>file:js\$ XMLHttpRequest file:include/ ioctl</code> <code>file:/usr/sys/kan/slp.c "You are not expected to understand this."</code>
<code>package:regexp</code>	Buscar paquetes cuyos nombres coincidan con <code>regexp</code> El nombre de un paquete es su URL o la información del servidor CVS. <code>package:perl "tar.gz Frodo" package:linux-2.6 int printf</code>
<code>lang:regexp</code>	Buscar sólo programas escritos en lenguajes que coincidan con <code>regexp</code> <code>lang:isp xml lang:"c++" sprintf "%s"</code>
<code>license:regexp</code>	Buscar sólo los archivos con licencia que coincidan con <code>regexp</code>

# Buscadores especializados

- ▣ Textos especializados y académicos → Digital.CSIC Acceso Abierto a Documentos Digitales <http://digital.csic.es/>
- ▣ Textos especializados y académicos → CiteSeer
- ▣ Textos especializados y académicos → Google Scholar
- ▣ Videos → YouTube
- ▣ Imágenes

## SCIRUS: For scientific information only

- ❑ **Scirus** es un motor de búsqueda específico de contenido científico. Semejante a [CiteSeer](#) y [Google Scholar](#), está enfocado a la información científica. A diferencia de CiteSeer, Scirus no se centra sólo en ciencias informáticas y no todo su contenido es de libre acceso (p.e. algunos resultados pueden estar en [PubMed](#) en cualquier revista de [Elsevier](#), requiriendo suscripción para su acceso. [www.scirus.com](http://www.scirus.com)

# Recuperación de Imágenes

- ❑ Tradicionalmente con metadatos, texto asociado a la imagen por nombre del fichero, texto de la página, etc
- ❑ Con la Web social con descriptores de los usuarios, p.e. Flickr

Otros buscadores de este tipo son:

- ❑ [Pixsy](#) imágenes y videos de Buzznet, flickr, iStockphoto, Fotolia, YouTube entre otros
- ❑ [Riya](#) buscador que incluye a Google, Yahoo, MSN, y flickr.
- ❑ [Tiltomo](#) basado en flickr
- ❑ [Liveplasma](#) para buscar música y video, busca por artista, grupo, película, director o actor.
- ❑ [Vdoogle](#) para buscar videos


# Recuperación de Imágenes

- ❑ Para buscar imágenes similares se intenta actualmente utilizar inteligencia artificial, para ello se necesita una base de datos de imágenes y descriptores asociados, como experimente el más impresionante es <http://images.google.com/imagelabeler/>
- ❑ Existen buscadores para buscar este tipo de material como [Retrievr](#) capaz de buscar a partir de una imagen (lo hace con la transformación [wavelet](#) en vez de una red de neuronas). Quizás sea el buscador más prometedor de este tipo



# Indicación por competición: Imágenes


- ❑ Flickr
- ❑ <http://www.espgame.org/>
- ❑ Google Image Labeler

Image LabelerGoogle Image Labeler

[Help](#) | [Sign In](#)

Welcome to **Google Image Labeler**, a new feature of Google Image Search that allows you to label random images to help improve the quality of Google's image search results.

[Begin as a guest](#)



### How does it work?

You'll be randomly paired with a partner who's online and using the feature. Over a 90-second period, you and your partner will be shown the same set of images and asked to provide as many labels as possible to describe each image you see. When your label matches your partner's label, you'll earn some points and move on to the next image until time runs out. After time expires, you can explore the images you've seen and the websites where those images were found. And we'll show you the points you've earned throughout the session.

For example, this image can be described by the labels: **bird**, **sky**, **soaring**, or **frigate bird**.

### What is required to participate?

Just an interest in helping Google improve the relevance of image search results for users like yourself. Although you do not have to log in to your Google account to help, logging in will allow you to keep track of your points. You can also choose to provide a nickname, or you can remain anonymous.

### Today's Top Pairs

1. lakagirl - guest	1300
2. guest - guest	1200
3. guest - guest	1200
4. guest - guest	1200
5. guest - Tomas LTU	1100

### All-time Top Contributors

1. wordgirl	1326400
2. tuki makeeta	1107700
3. FrankD	997400
4. 15minL8	986300
5. Nicio	965200

# Buscar en repositorios de bibliotecas

European Library  
OpenDOAR  
ROAR  
WorldCat (OCLC)

The image displays four different digital library search interfaces:

- The European Library:** A screenshot of the website in a Mozilla Firefox browser. It features a search bar, navigation tabs (SEARCH, COLLECTIONS, TREASURES, LIBRARIES), and a list of search results with columns for Name, Date, and Thumbnail.
- OpenDOAR:** A screenshot of the OpenDOAR website, showing a search bar and a list of search results. The text "The European Library is for searching the content of European national libraries." is visible.
- Collection List:** A screenshot of a collection list with a sidebar menu containing categories like "posters", "photographs", "portraits", "music", "children's literature", "manuscripts", "digitized books", "maps & atlases, cartography", "newspapers", and "scientific articles".
- WorldCat:** A screenshot of the WorldCat website, showing a search bar and a list of search results. The text "Encontrar en una biblioteca" is visible.

# Búsqueda en Bibliotecas

## ❑ Z39.50

<http://www.csic.es/cbic/catalogosz.html>

<http://www.loc.gov/z3950/gateway.html>

## ❑ BookSearch

## ❑ Bibliotext

## ❑ ProCite

The screenshot displays the BookWhere Online web application. The main window shows a search results table with columns for Author and Title. The table lists several records, including works by Viñola Morató, Manuel Simón, Morató, Emica, Reyes Huertas, Antonio, Morató y Aragónés, Josep María, Uruguay, and Opónsala. The interface includes navigation tabs for 'My Summary', 'My Records', 'Simple', and 'Precision'. Below the table, there are links for 'Export Now' and 'Close'. The bottom of the screen shows a status bar indicating 'Viewing 1 to 20 of 50 expected records' and a 'Summary in Italian' link.

BookWhere Online (Build 1.0.70060811, 1) - Microsoft Internet Explorer

Export Records

BookWhere Online provides users with the ability to both download selected records directly from this page, or have the user e-mail the records to a provided e-mail address.

Step 1: Confirm Export Format

BookWhere Online Supports a number of different Export Formats.

Please confirm the selected export format is correct:

Step 2: Send the Exported Records

BookWhere Online provides a number of options to send the exported records to you. You can view them as text in a new browser window, save them to a file, or even e-mail the records to any given e-mail address. Please select the appropriate method and then click the **Export Now** button to send these exported records to you.

Viewing 1 to 20 of 50 expected records

export save selected status info: new search help next

Summary in Italian

Number of Pages: 114 p.

