

CLASE II

Investigación en Internet. Búsquedas de información en la Web Invisible y Semántica



CREATIVE COMMONS RECONOCIMIENTO-NOCOMERCIAL-COMPARTIRIGUAL
3.0 UNPORTED LICENSE



Las WWW's...

En la clase anterior se introdujo el concepto de WWW y se mencionó su clasificación (Visible, Invisible y Semántica).

Focalizando en la **WEB VISIBLE**, se presentaron especialmente dos tipologías de buscadores existentes en esta Web:

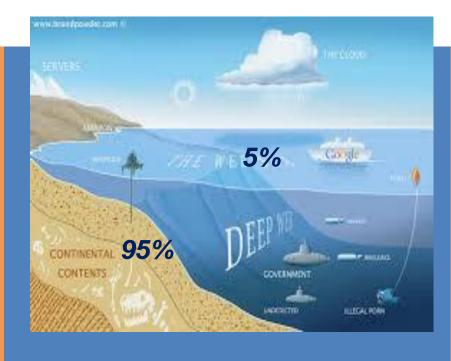
- 1) Directorio (centrado en la búsqueda por temas o categorías)
- 2) Motor de búsqueda (centrado en la búsqueda por palabra/frase clave)

En esta clase continuamos con las búsquedas de información centrándonos en la Web Invisible, analizando el desarrollo de la Web Semántica y finalizando con una reflexión respecto del proceso de búsqueda en estas Webs.



Web Invisible o Profunda u Oculta...

Se define como la "porción de la Web que no puede ser indexada por los buscadores comunes". Se compone del 95% de la información existente en la WWW, contra el 5% de la Web Visible...



En los distintos niveles de las aguas profundas se encuentra información, oculta, utilizable en algunos casos a nivel profesional y académico.





La **WEB INVISIBLE** contiene información que **no puede ni debe ser visibilizada por un buscador común**, y en consecuencia no puede ser indizada por esta herramienta.

Esta información está contenida en diferentes Webs, entre las que podemos mencionar:

- La Web Opaca
- ·La Web Privada
- La Web Propietaria y
- •La Web Realmente Invisible...



Web Invisible o Profunda... WEB OPACA ⇒ Se compone de contenidos que

1

no son indexados por un buscador por cuestiones de economía.



En este primer ejemplo, Google devuelve solo dos resultados del sitio web: Flickr...

A. Fernández: "Cristina va a cambiar si se le antoja" - Río Negro www.rionegro.com.ar/.../anibal-fernandez-cristina-va-a-cambiar-si-se-le-... ▼ El proyecto no cambia sostuvo el senador al analizar la derrota en las elecciones primarias Aseguro que el secretario de Comercio Moreno no toma decis.

Cleri: "Massa es una puesta en escena de aquellos que nos quieren ... www.lacapital.com.ar/.../Cleri-Massa-es-una-puesta-en-escena-de-aquello... ▼ hace 20 horas - A su vez, reconoció que después de los resultados en las elecciones primarias, la presidenta se mostró "dispuesta" a modificar lo que sea ...

Se cargaron cerca de 10 mil kilos de papel y una importante ...

www.elcomercial.com.ar/index.php?...id... •

hace 14 horas - Se juntaron cerca de diez mil kilos de papel y cartones, entre ellos urnas y votos de las últimas **elecciones primarias**, además de una ...

Eduardo Mendoza, Mauricio o las Elecciones Primarias | Flickr ... www.flickr.com/photos/34782241@N00/196239501/ * 23/07/2006 - Eduardo Mendoza, Mauricio o las Elecciones Primarias.

Debates y elecciones primarias | Flickr - Photo Sharing!

www.flickr.com/photos/maicopupsa/5384156618/ -

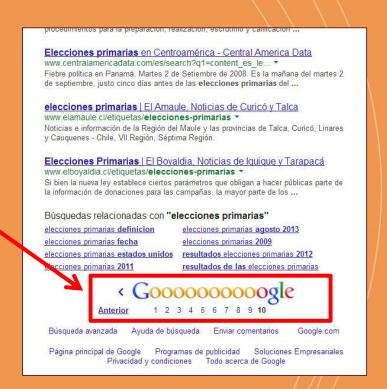
14/01/2011 - Los alumnos de MAICOP IV se enfrentan a sus compañeros en los debates electorales y las votaciones primarias.



En este segundo ejemplo, Google solo deja acceder a 100 de sus resultados...



Solo muestra 100 resultados aunque señala disponer de más de 2 millones...





2

contenidos que no son indexados dada la frecuencia de indización del buscador...

Último rastreo: 30/07 Próximo rastreo: 30/08





Las páginas creadas del 31/07 al 29/08 no aparecerán como parte de los resultados...

Estas imágenes intentan señalar que todos los sitios web creados después del 30/07 y antes del 30/8 no aparecerán como resultados de una búsqueda, hasta que se produzca un nuevo rastreo, una nueva actualización por parte del buscador...



3

contenidos con bajo número de Page Rank,

lo que hace que el buscador no lo muestre como parte de sus resultados



Solo se muestran las páginas con mayor Page Rank...

procedimientos para la preparación, realización, escrutinio y callicación ...

Elecciones primarias en Centroamérica - Central America Data

www.centralamericadata.com/es/search?q1=content_es_le... •

Fiebre política en Panamá. Martes 2 de Setiembre de 2008. Es la mañana del martes 2 de septiembre, justo cinco días antes de las **elecciones primarias** del ...

elecciones primarias | El Amaule, Noticias de Curicó y Talca

www.elamaule.cl/etiquetas/elecciones-primarias *

Noticias e información de la Región del Maule y las provincias de Talca, Curicó, Linares y Cauquenes - Chile, VII Región, Séptima Región.

Elecciones Primarias | El Boyaldia, Noticias de Iquique y Tarapacá www.elboyaldia.cl/etiquetas/elecciones-primarias ▼

Si bien la nueva ley establece ciertos parámetros que obligan a hacer públicas parte de la información de donaciones para las campañas. la mayor parte de los ...

Búsquedas relacionadas con "elecciones primarias"

elecciones primarias definicion elecciones primarias agosto 2013

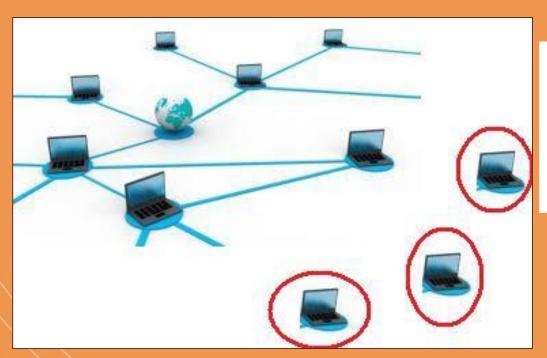
elecciones primarias fecha elecciones primarias 2009

elecciones primarias estados unidos resultados elecciones primarias 2012

resultados de las elecciones primarias



Página principal de Google Programas de publicidad Soluciones Empresariales Privacidad y condiciones Todo acerca de Google Existen muchas razones por las que una página puede tener un bajo posicionamiento en las búsquedas: pocas visitas, poca cantidad de enlaces, poca autoridad de quienes escriben en el sitio, etc. Esto ocasiona entonces que las páginas queden fuera del rastreo: "poco conectadas" o directamente "desconectadas"...



Quedan entonces fuera del rastreo!!!



Web Invisible o Profunda...

WEB PRIVADA

Contiene información confidencial, jamás accesible para el buscador. Se compone de

páginas protegidas por contraseñas

Carpetas

Última actualización: Mie, 1:49 pm (Comprobar correo)

ENTRADA (1501) INBOX Drafts

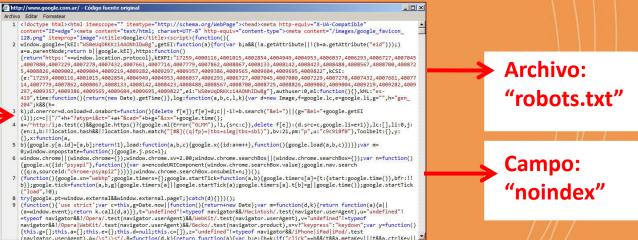
INBOX Sent INBOX Trash Consultas 2008 Cursos Info Importante Info para alumnos

| Carpeta actual: ENTRADA | | <u>Desconectarse</u> |
|---|----------------------|--|
| Componer Direcciones Carpetas Opciones Busc | <u>ar Ayuda Spam</u> | Filter SquirrelMai |
| Anterior Siguiente 1 2 Todos Cambia todos | | Viendo mensajes: del 1 al 1000 (total 1501) |
| Mover seleccionados a: ENTRADA Mover Reenviar | | Marcar mensajes seleccionados como: Leído No leído Borrar |
| De 🗉 | Fecha 📤 | Asunto 🗆 |
| ☐ La Carpinteria | 20/11/2012 | ⁺ Reservá tu entrada para el céntesimo mono de Osqui Guzmá |
| ☐ Palazzo Diego | 8:21 pm | Filtro "bajo mesada" 3M Cuno |
| ☐ Ricardo Aguilar | 8:00 pm | Promocion Dispenser de Agua! |
| ☐ Piero Alvaro | 5:08 pm | Ofertas en Dispenser de Agua |
| ☐ Quesada Cristian | 4:45 pm | El mejor filtro "Bajos Mesada" 3M |
| ☐ Lopez Marcos | 4:25 pm | Promocion Dispenser de Agua! |
| ☐ Piero Elena | 2:21 pm | Dispenser de Agua |
| ☐ Sabri Judith | 1:26 pm | Dispenser de Agua |
| ☐ Guerrero Cora | 1:09 pm | No gaste mas en bidones de agua! |
| ☐ Prensa UNSL | 1:08 pm | + Noticias UNSL- 31 de Octubre de 2012 |
| ☐ Fundación Civis | 12:50 pm | + No pagues de mas |
| ☐ Sabri Ricardo | 12:11 pm | Dispenser de Agua |

páginas que expresamente solicitan no ser indexadas

Existen archivos especiales o campos específicos asociados/dentro de una página web que explicitan que la página no sea indexada...



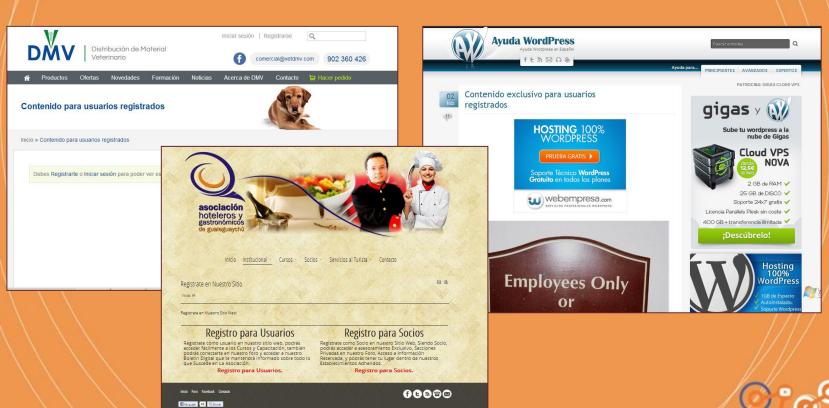






Web Invisible o Profunda...

WEB PROPIETARIA Contiene páginas en las que es necesario registrarse para acceder al contenido...



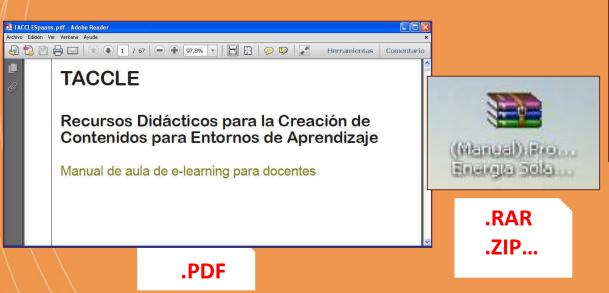
0



Web Invisible o Profunda...

WEB REALMENTE INVISIBLE | incluye

archivos de la Web en diversos formatos: .pdf, .js, .ppt, etc.









Afortunadamente, muchos buscadores han evolucionado y permiten encontrar diversidad de tipos de archivos...



b

páginas que se generan dinámicamente, esto es, en tiempo de ejecución (por ej. al realizar una consulta en una base de datos).

Página estática Búsqueda Bibliográfica Tipo de Material Criterio Búsqueda libre Término(s) a buscar: Buscar

Página dinámica



C

páginas que se generan en tiempo real (por ej. ante la transmisión en vivo y en directo de un partido).

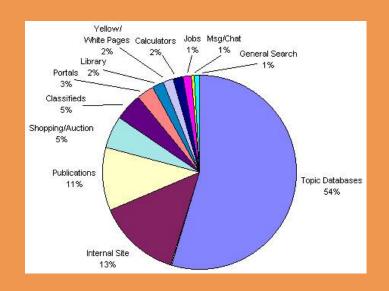
Página generada en tiempo real





Web Invisible o Profunda...

¿CÓMO ACCEDEMOS A ELLA?



A través de buscadores especializados en bases de datos que trabajan generalmente con las tipologías: directorio y/o motor de búsqueda.



BUSCADORESWeb Invisible o Profunda...

1

Motor de Búsqueda GOOGLE (en su clasificación general/avanzada)

leda avanzada

| / | Buscar páginas con | | | Haz lo siguiente en el cuadro de búsqueda |
|---|--|---------------------------------|------------------|---|
| | todas estas palabras: | I | | Escribe las palabras importantes: terrier ratonero tricolor |
| | esta palabra o frase exactas: | | | Escribe las palabras exactas entre comillas: "terrier ratonero" |
| | cualquiera de estas palabras: | | | Escribe OR entre todas las palabras que quieras: miniatura OR estándar |
| | ninguna de estas palabras: | | | Añade un signo menos delante de las palabras que no quieras que aparezcan: -roedor, -"Jack Russell" |
| | números desde el: | hasta | | Escribe dos puntos seguidos entre los números y añade una unidad de medida: 1035 kg, 300500 euros, 20102011 |
| | | | | |
| | A continuación, limitar los resultados por | | | |
| | idioma: | cualquier idioma | * | Busca páginas en el idioma que selecciones. |
| | región: | cualquier región | * | Busca páginas publicadas en una región determinada. |
| | última actualización: | en cualquier momento | * | Busca páginas actualizadas en el periodo de tiempo especificado. |
| | sitio o dominio: | | | Busca un sitio (como wikipedia.org) o limita los resultados a un dominio como, por ejemplo, .edu, .org o .gov. |
| | los términos que aparecen: | en cualquier lugar de la pàgina | * | Busca términos en toda la página, en el título de la página o en la dirección web, o enlaces a la página que estás buscando. |
| | Búsqueda Segura: | 11 | * | Indica a Búsqueda Segura si quieres que filtre contenido sexualmente explícito. |
| | tipo de archivo: | cualquier formato | Ţ | 5a páginas en el formato que prefieras. |
| | derechos de uso: | sin flitrar por licencia | ٧ | Busca páginas que puedas utilizar libremente. |
| | | В | Dequeda avanzada | |

Opciones de restricción de resultados

Operador: FILETYPE

También puedes...

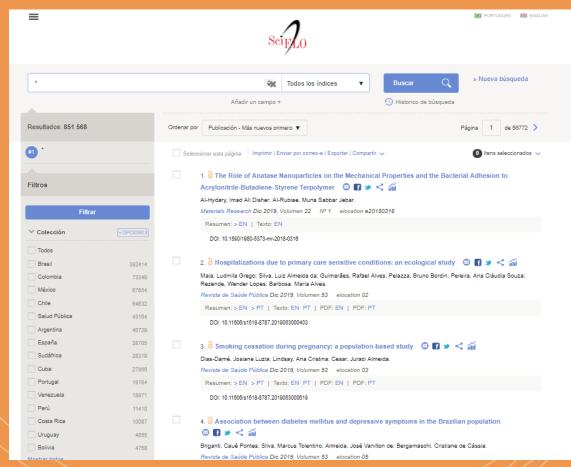
Suscar páginas similares a una URL o crear un enlace a la misma scar páginas visitadas peradores en el cuadro de búsqueda ar configuración de búsqueda Más opciones de restricción de resultados



BUSCADORES

Web Invisible o Profunda...

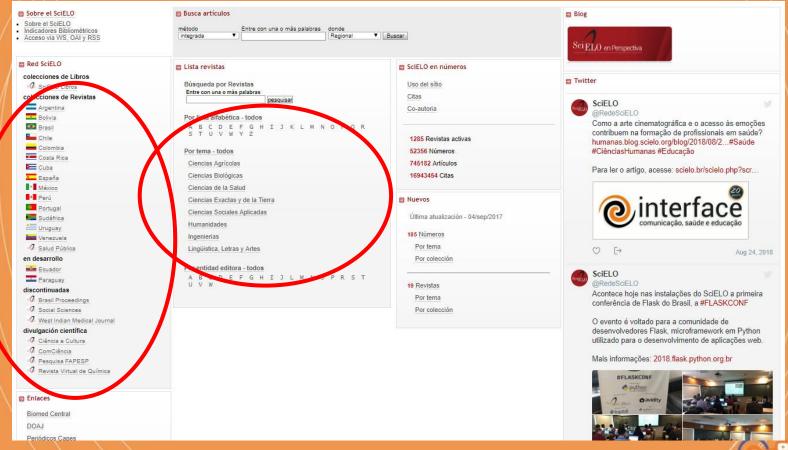
Directorio/ Motor de Búsqueda SCIELO (http://www.scielo.org/php/index.php?lang=es)



Biblioteca científica electrónica en línea...



Búsqueda por Categorías



Países...



Búsqueda por Categorías

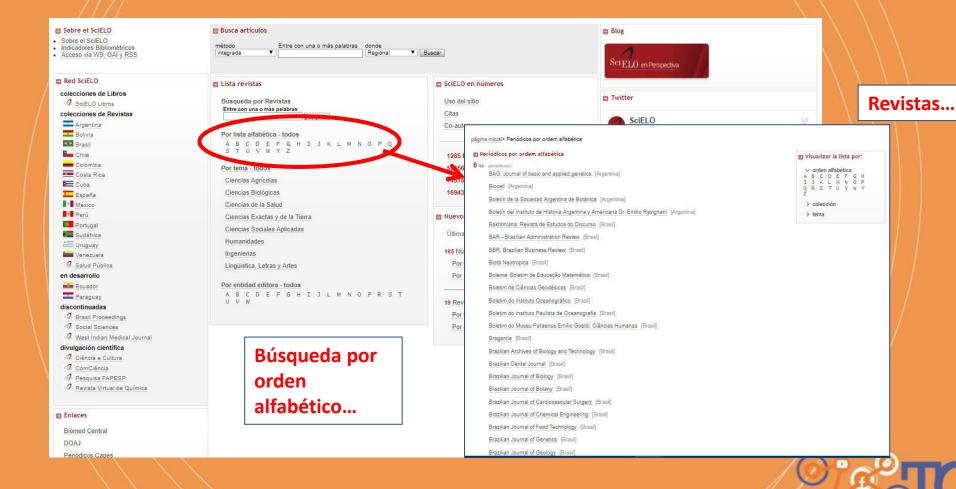


Países...

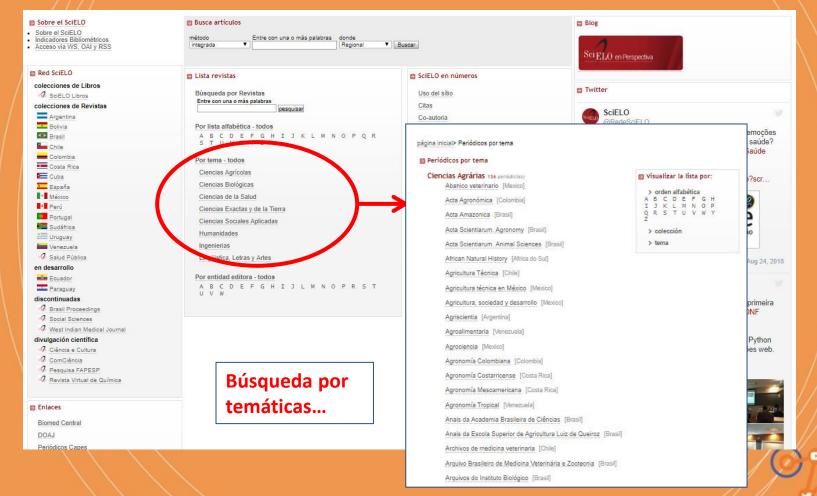
- Enfoques revista de la Universidad Adventista del Pla
 Espacios en blanco. Serie indagaciones 16 números
- Estudios Centro de Estudios Avanzados, Universidad Nacional de Córdoba 17 númer



Búsqueda por Categorías



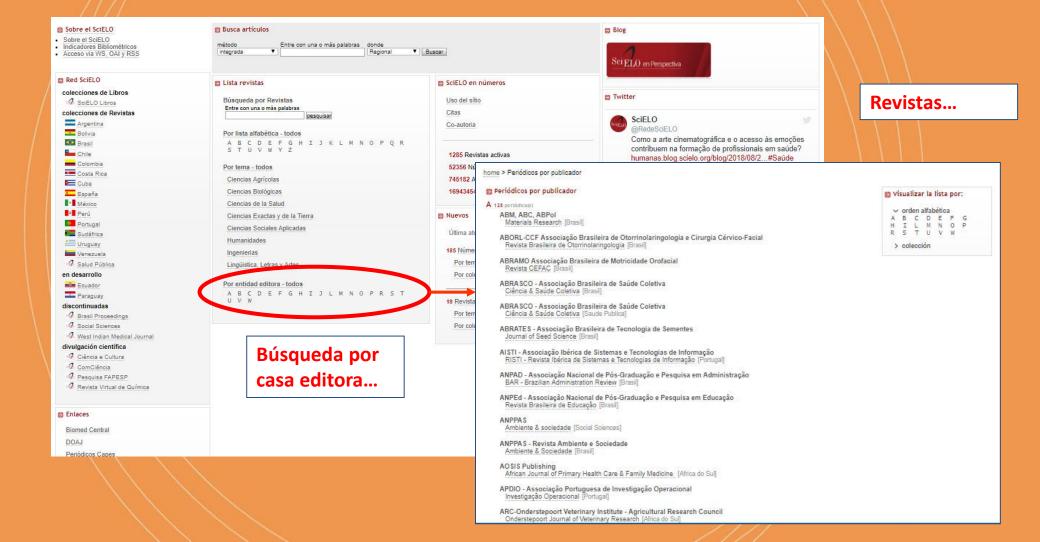
Búsqueda por Categorías



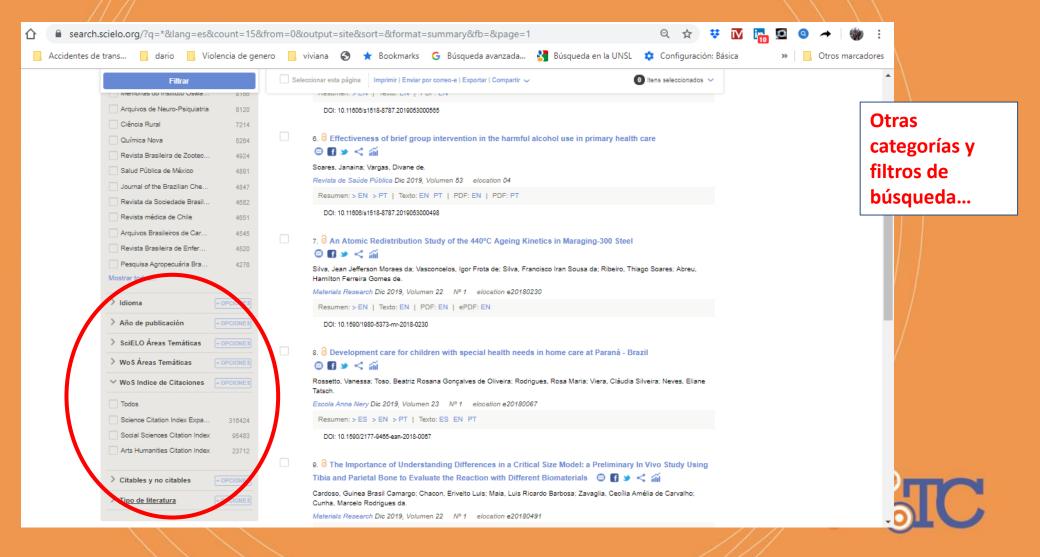
Revistas...



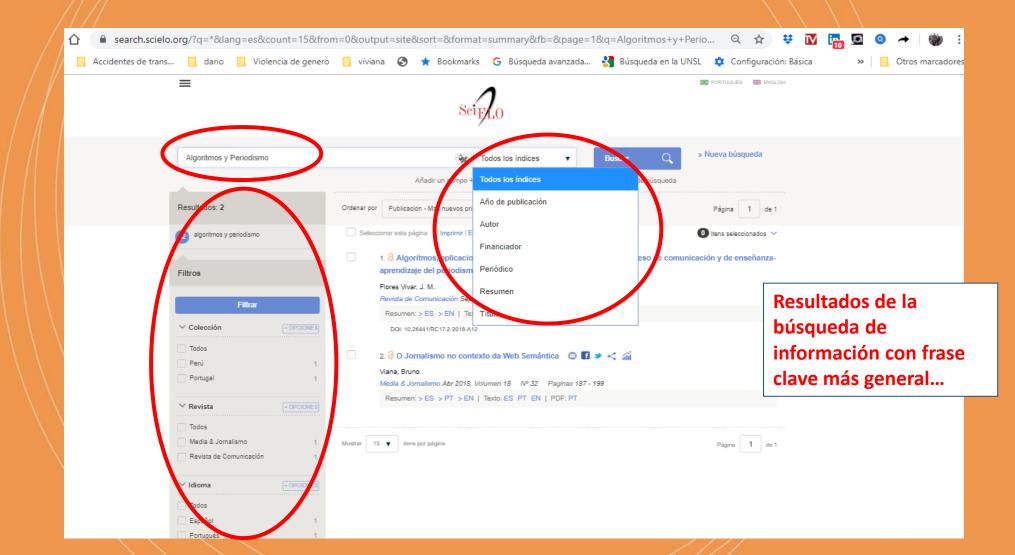
Búsqueda por Categorías



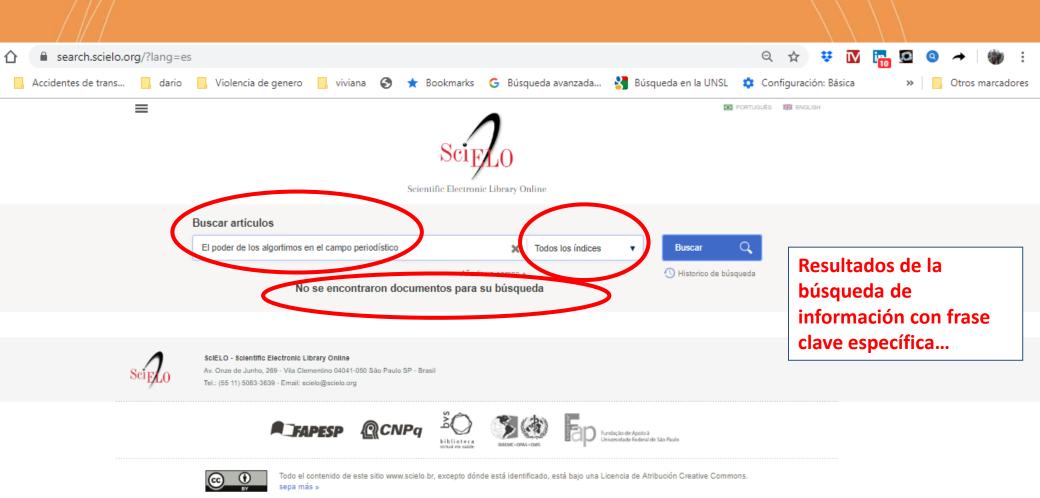
Búsqueda por Categorías



Búsqueda por Palabra/Frase clave



Búsqueda por Palabra/Frase clave



BUSCADORES

Web Invisible o Profunda...

3

Motor de búsqueda Generalista
WAYBACK MACHINE (http://archive.org/web/web.php)



Permite el acceso a versiones de sitios web que pueden o no estar en línea...



Explore more than 376 billion web pages saved over time

http://

Find the Wayback Machine useful?















BROWSE HISTORY







Wayback Machine Availability API Build your own tools.

WordPress Broken Link Checker Banish broken links from your blog.

404 Handler for Webmasters

Help users get where they were going.



Subscription Service

Archive-It enables you to capture, manage and search collections of digital content without any technical expertise or hosting facilities. Visit Archive-It to build and browse the collections



Save Page Now

http://

SAVE PAGE

Capture a web page as it appears now for use as a trusted citation in the future.

Only available for sites that allow crawlers.

WAYBACK MACHINE

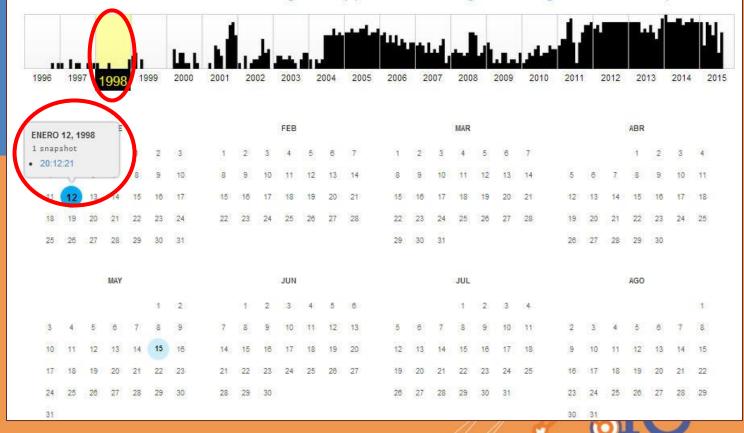
Para buscar se coloca la dirección web del sitio y luego se selecciona el año, mes y día que se desea visualizar (siempre que esté disponible)...



http://www.clarin.com

Saved 5.763 times between octubre 31, 1996 and agosto 24, 2015.

PLEASE DONATE TODAY. Your generosity preserves knowledge for future generations. Thank you.



WAYBACK MACHINE

El resultado es la captura de la versión en línea de esa información, con la posibilidad de ingreso a su página online...





BUSCADORES

Web Invisible o Profunda...

Motor de búsqueda Generalista **SEARCH**(https://search.creativecommons.org)

Busca contenidos (audio, imagen, video, etc.) con licencia **Creative** Commons... **About Browse by Collection Feedback** Looking for the old CC Search portal? Go here creative commons Search for images... Q Use for commercial purposes
Modify or adapt Search for free content in the public domain and under Creative Commons licenses. Learn more about CC licenses here.



(cc)Search

We'd love to hear from you!

See our Search Syntax Guide here (?)

Otros ejemplos de buscadores:

Bases de datos científicas y/o académicas:

- Base
- Dialnet
- SciCentral, Latindex, Doaj, Worldwidescience, etc.

Recursos en diversos formatos:

- World Cat, U.S. Geological Survey, etc.

Recursos confidenciales:

- The National Security Archive
- CIA, etc.

Bases de datos con recursos y/o literatura en medicina y/o salud:

- Lilacs
- Cuiden
- ScienceDirect
- PubMed, etc.



BUSCADORES

Web Invisible o Profunda...

Poseen una base de datos propia (elaborada por personas o robots) compuesta por accesos a sitios web confiables, aunque puedan estar desactualizados.

Permiten la configuración de las bases de datos donde buscar, la tipología de archivos y el lugar y cantidad de resultados que devuelven.

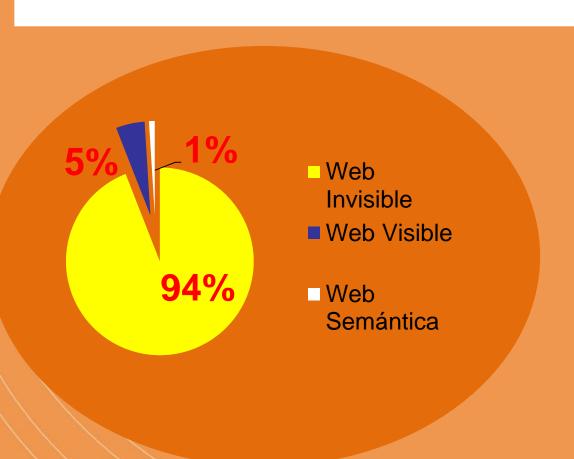
¿CÓMO TRABAJAN LOS BUSCADORES PROFUNDOS?

Se utilizan para buscar información especializada en torno a una temática, de carácter científico.

Poseen opciones de búsqueda por categorías o palabra clave. En este último caso permiten especificar: título del recurso, autor del recurso, descripción o resumen del recurso, contenido del recurso y palabras clave, que permiten identificarlo, entre otros.

WWW Semántica

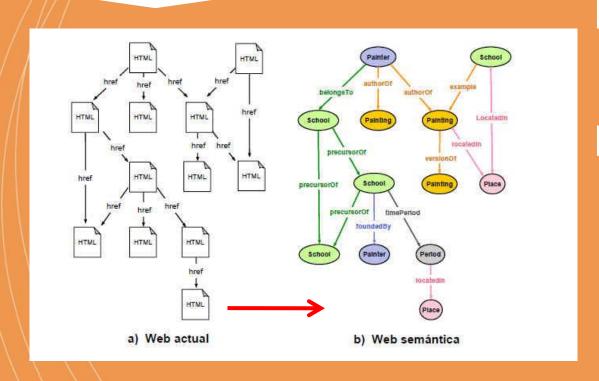
Se define como "...una extensión de la Web existente que permite la realización de búsquedas más inteligentes a través de buscadores inteligentes".



Su objetivo es mejorar las búsquedas, de forma tal que las máquinas no solo sean capaces de utilizar la información, sino también interpretarla, asemejándose lo más posible al ser humano...

WWW Semántica

"Quiero viajar a Lima, Perú, y sólo quiero gastar 5000 dólares"



REQUIERE:

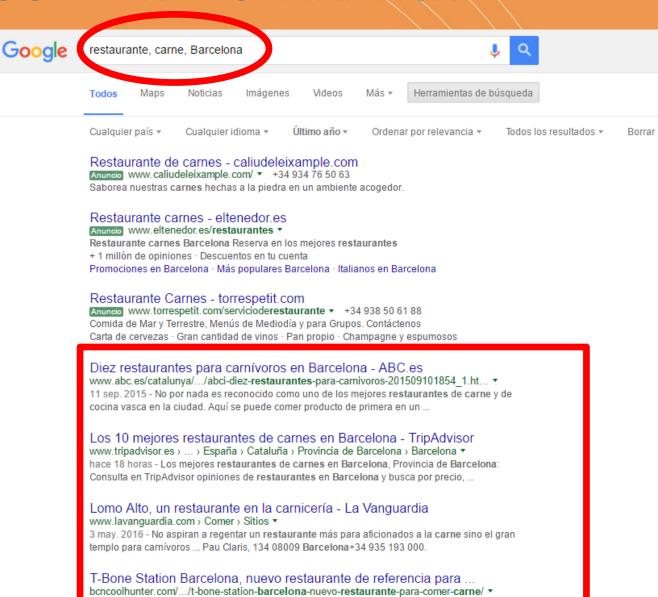
- •OWL: permite definir conceptos u ontologías que puedan ser utilizadas por usuarios, sistemas y bases de datos.
- •RDF: describe conceptos y ontologías y cómo se encuentran relacionados.
- •SPARQL: lenguaje de consulta sobre las ontologías que permite hacer búsquedas sobre los recursos de la Web

La reestructuración de la Web podría funcionar de manera potente pero se proyectan muchos años para su óptimo funcionamiento...



GOOGLE Y WEB SEMÁNTICA

Y aunque se trabaja para lograr este tipo de búsquedas aún está en proceso de desarrollo ya que no siempre las búsquedas funcionan de manera inteligente...



22 mar. 2016 - Para los amantes de la carne, Barcelona cuenta con un nuevo sitio de referencia, el

T-Bone Station, El restaurante, que acaba de inaugurar en el número 14.

FIN DE PRESENTACIÓN Gracias por su atención...

Elaborado por Esp. Viviana M. Ponce
vivianaponce@gmail.com
Diseño gráfico: DIG Luisina Andreoni
andreoniluisina@gmail.com
Fac. de Cs. Humanas- UNSL

