

BIBLIOTECA UNIVERSITARIA

Fuentes de información en internet

Material formativo



Reconocimiento – NoComercial-CompartirIgual (By-ns-sa): No se permite un uso comercial de la obra original ni de las posibles obras derivadas, la distribución de las cuales se debe hacer con una licencia igual a la que regula la obra original.

ÍNDICE

Introducción	2
Directorios	2
Repositorios	4
Motores de búsqueda	8
Metabuscadores	11

FUENTES DE INFORMACIÓN EN INTERNET

Introducción

La recuperación de la información en Internet tiene un carácter selectivo y organizado. Tiene unas características diferentes a la localización de información tradicional en un centro de documentación o biblioteca.

Básicamente se basa en dos paradigmas, aunque su esquema es idéntico: de un conjunto se selecciona unos ítems organizados por la semejanza a una pregunta.

Ambos sistemas se enfrentan al mismo problema, encontrar el escenario ideal de recuperación, pero desde perspectivas diferentes, encontrar el mayor número de documentos recuperados y relevantes:



BROWSING o navegación: consiste en elegir sucesivamente, de manera no sistemática, por medio de una determinada estructura de información preestablecida, como un cuadro de clasificación o una estructura de categorías. Es el tipo de recuperación de la información que fundamenta los directorios de Internet. Se navega eligiendo una categoría y sucesivas subcategorías hasta que se localiza lo que se busca.



QUERYING o interrogación: está basado en los motores de búsqueda, que posibilitan la recuperación de información a partir de palabras clave gracias a una base de datos que indiza la información recibida por medio de un programa robot que recorre la Red.

Directorios



Los directorios consisten en una selección de recursos web organizados según una estructura jerárquica de materias (desde categorías más amplias a categorías más específicas).

Búsqueda

La búsqueda en los directorios la hacemos mediante navegación (browsing). Comenzaremos desde categorías generales e iremos descendiendo hasta las materias más específicas que sean de nuestro interés.



Ten en cuenta que cada directorio utiliza sus propias categorías o materias

Contenido

En los directorios NO se encuentra toda la información o todos los recursos que hay sobre un tema en la web.

A diferencia de los motores de búsqueda, los directorios contienen un conjunto limitado de recursos que han sido previamente analizados, seleccionados y clasificados de forma manual o humana (no realizado por robots), y que garantiza la pertinencia y calidad de los contenidos.



Selecciona aquellos directorios que sean actualizados frecuentemente y que informen de los criterios de selección utilizados.

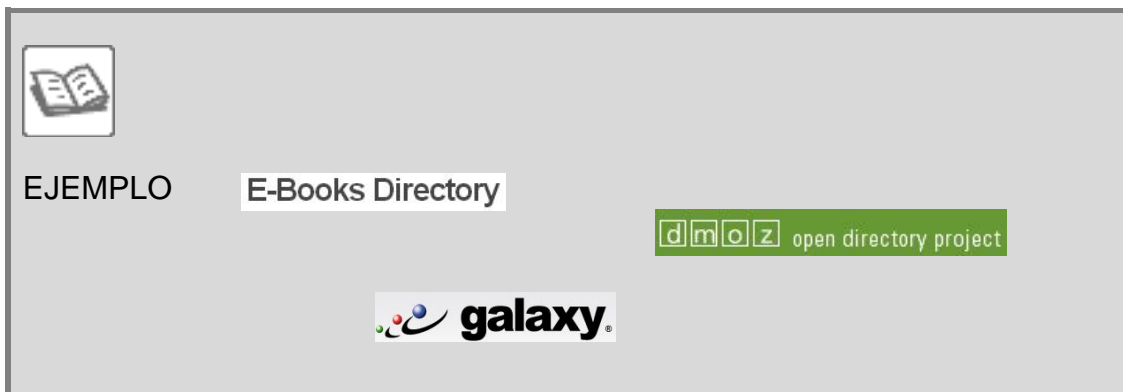


Usa los directorios para buscar bien información general de un tema, bien información específica que pertenezca de manera clara a una categoría.

Clasificación

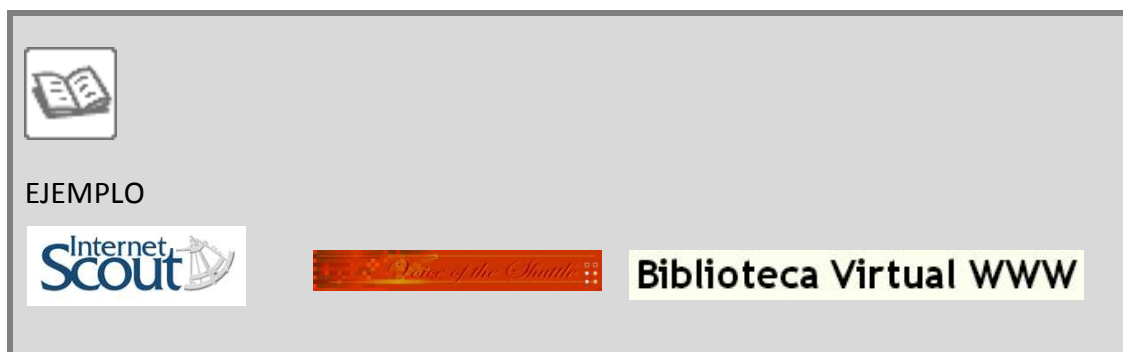
Los directorios podemos clasificarlos en:

- **Generales:** Ofrecen información de carácter general en cualquier tipo de ámbito: deportes, entretenimiento, tecnología...: *E-Books Directory, Galaxy, Open Directory Project, Academic Info,...*



■ **Académicos:** Contienen información científica o dirigida al mundo de la docencia e investigación. Pueden ser:

- Temáticos: [Guías temáticas de la BUA](#), Voice of the Shuttle (arte y humanidades)...
- Multidisciplinares: Biblioteca Virtual WWW, PINAKEΣ...



Para más información sobre las Guías temáticas de la BUA consulta el tema 4.1.8

Repositorios

Definición y características

Según la R.A.E., repositorio es un "lugar donde se guarda algo".



En nuestro ámbito, entendemos repositorio como un conjunto de documentos almacenados en formato electrónico a los que podemos acceder en línea con las siguientes características generales:

- Acceso libre o abierto
- Punto de acceso único a los documentos mediante el uso de identificadores persistentes (evita enlaces rotos, obsoletos...)
- Descarga de texto completo o del archivo íntegro (salvo excepciones)
- Uso de metadatos con protocolo OAI-PMH (Open Archives Initiative-Protocol Metadata Harvesting)



Usa los repositorios para tus búsquedas. Son de acceso libre, tienes gran cantidad y diversidad de documentos y hay mucha información en español.

Búsqueda

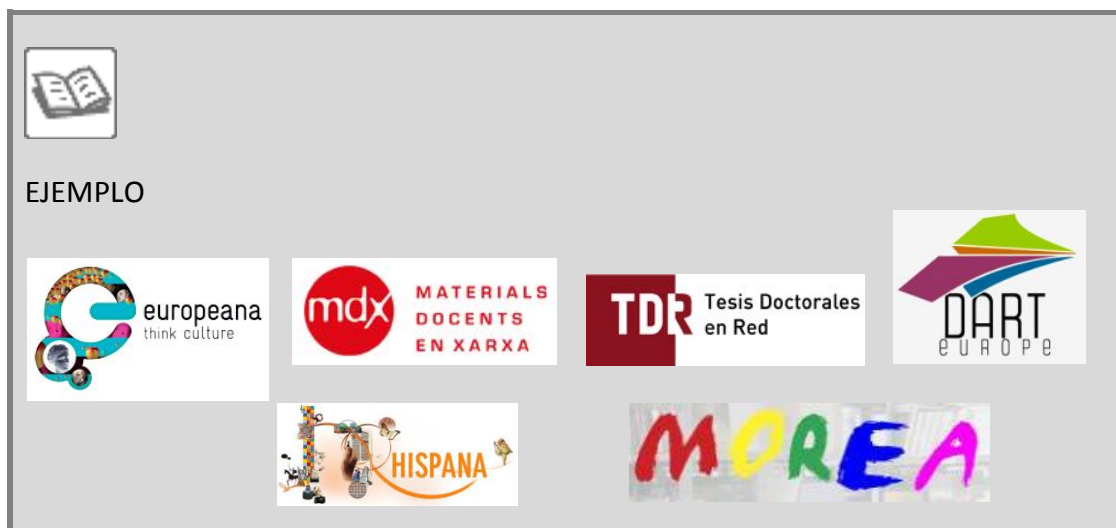
Las búsquedas en los repositorios combinan los sistemas de navegación (browsing) y el uso de motores de búsqueda (querying)

Contenido

El contenido de los repositorios dependerá del ámbito de cada uno de ellos. Puedes encontrar gran diversidad de documentos, entre ellos:

- **Artículos:** Son el tipo de documento con mayor presencia en los repositorios
- **Libros:** Hispana...
- **Tesis:** TDX, DART-Europe E-theses Portal...
- **Fotografías, audiovisuales, multimedia:** Europeana...

- **Materiales docentes** (temarios, prácticas, exámenes...): MOREA, MDX...



Clasificación

Los repositorios se pueden clasificar en las siguientes categorías:

- **Institucionales:** Recoge la documentación generada por una institución en el desarrollo de su actividad institucional, tales como publicaciones del personal docente e investigador, tesis, material docente, congresos...: *RUA, Digital.CSIC...*
- **Temáticos:** Recogen documentos especializados por la materia o área temática: *PubMed Central* (ciencias de la salud), *arXiv.org* (física, matemáticas...)
- **Agregadores o Recolectores:** Agrupan múltiples repositorios permitiendo consultarlos simultáneamente mediante un único formulario de búsqueda: *Recolecta* (repositorios científicos nacionales), *OpenAIRE* (investigación europea), *Scientific Commons*, ...



Para más información sobre RUA consulta el tema 4.1.9

Recolecta



Recolector de Ciencia abierta que agrupa a todos los repositorios científicos nacionales en acceso abierto, permitiendo el acceso libre a toda la producción científica depositada en ellos.

Los objetivos de [Recolecta](#) son, entre otros:

- Promover el acceso abierto en la investigación nacional
- Dar mayor visibilidad nacional e internacional a los resultados de la investigación realizada en España.

Los resultados de la búsqueda en [Recolecta](#) provienen de los diferentes repositorios que agrupa, por lo que puedes encontrar gran variedad de documentos como:

- Tesis doctorales
- Comunicaciones a congresos
- Artículos científicos
- Trabajos de investigación
- Materiales docentes y objetos de aprendizaje

GOBIERNO DE ESPAÑA
MINISTERIO DE ECONOMÍA Y COMPETITIVIDAD

FECYT
FUNDACIÓN ESPAÑOLA PARA LA CIENCIA Y LA TECNOLOGÍA

Recolecta
RECOLECTOR DE CIENCIA ABIERTA

BUSCADOR DE CIENCIA ABIERTA

BÚSQUEDA SIMPLE | BÚSQUEDA AVANZADA | NAVEGAR POR RECURSOS

Buscar en más de **830.000** documentos en **81** recursos



Directorios de repositorios

Para conocer y obtener información de los distintos repositorios en acceso abierto que hay disponibles, puedes utilizar los directorios de repositorios, entre ellos:

- **BuscaRepositorios:** Repositorios Institucionales Españoles de Acceso Abierto
- **OpenDOAR:** Repositorios académicos de acceso abierto
- **ROAR:** Registro de repositorios en acceso abierto

Motores de búsqueda (search engines)

Te posibilitan la recuperación de información en la web por medio de descriptores o palabras clave.



Son herramientas que utilizan unos programas que rastrean y recogen la información de Internet para incorporarla a una base de datos que se pueden consultar por medio de los buscadores

Todos tienen la misma estructura, con cuatro elementos fundamentales:

- **Spider o robot lógico:** Tiene la finalidad de rastrear de una forma constante la Red, recopilando información de todos los documentos a los que pueden acceder. Para desplazarse utilizan los enlaces de hipertexto que encuentran o directamente navegan a las URL para localizar a otros documentos en Internet.
- **Programa de indexación:** Elabora un índice inverso que permitirá la posterior recuperación de los documentos a partir de la representación o del documento en sí. Están formados mayoritariamente por palabras clave y su ubicación en el texto.

Actualmente hay intentos de elaborar una indexación por conceptos, de acuerdo con teorías lingüísticas y a la inteligencia artificial (AI).

- **Base de datos:** Contiene los elementos que son confrontados con los descriptores o palabras clave que utiliza el usuario para formular la pregunta. Generalmente es un índice inverso que contiene las palabras que aparecen en el texto, la frecuencia y la URL de todos los elementos indexados.
- **Interfaz del usuario:** Es la "pantalla", la parte visible del buscador que te permite la interacción con la base de datos. Generalmente se trata de un formulario donde se introduce la ecuación de búsqueda más o menos compleja con las palabras clave por las que se quiere recuperar los documentos que satisfagan las necesidades de información.

Las características principales de un motor de búsqueda son:

- Funcionan por interrogación
- La organización es automática, lo que puede inducir a errores
- No hacen ninguna discriminación, el robot incluye todo lo que lee
- El alineamiento es específico para cada motor. Cuando hablamos de alineamiento nos referimos al orden de aparición de los resultados
- Actualización frecuente, algunos incluso diaria
- Exhaustividad en los resultados, es decir te ofrecen gran cantidad de respuestas.
- Acceso a grandes volúmenes o colecciones de información



Como recomendación es importante que consideres la utilización de este recurso para cuando busques:

- Un concepto o una frase específica
- Un tema muy concreto del que tienes pocos datos
- Obtener la máxima información posible



También es importante que contrastes los resultados entre varios de ellos, ya que el resultado puede variar de uno a a otro.

Aunque estemos muy habituados al manejo de buscadores, especialmente por las facilidades que nos ofrece Google como herramienta más utilizada, no debes olvidar que tienen una serie de inconvenientes y tienes que analizar la información recuperada de una forma objetiva.

Las dos desventajas fundamentales son:

- La necesidad de conocer cómo interrogar el motor de búsqueda. Una sintaxis que puede variar de uno a otro
- El desbordamiento cognitivo que pueden producir por la ingente cantidad de información que presentan, que hace que sea difícil asimilarla y producir lo que se conoce como *infoxicación*.



EJEMPLO

Los buscadores más importantes que obtienen los datos por medio de motores de búsqueda son:

- AOL
- Ask
- Baidu
- Bing
- Blekko
- EntireWeb
- Google
- HotBot
- KwMap
- Lycos
- ScrubTheWeb
- Yahoo! Search

Metabuscadores

Existen otras herramientas que hacen uso de las bases de datos de los motores de búsqueda. Se trata de los METABUSCADORES.



Permiten realizar búsquedas en varios buscadores de manera simultánea, ahorrando tiempo y recursos a evitar tener que interrogar a cada uno de ellos por separado.

Los metabuscadores por medio de una única interfaz, recogen en un formulario las necesidades de información del usuario y, posteriormente la contrastan con la base de datos o índices de alguno de cada buscador con los que trabaje.



A pesar de las ventajas de este sistema y lo cómodo que puede llegar a ser, tienes que tener en cuenta que:

- No todos los motores aceptan el mismo lenguaje de interrogación, por lo que la sintaxis se reduce a los elementos más comunes y hace que no puedas explotar el potencial de cada motor de búsqueda como harías por separado,
- Los más avanzados posibilitan el control de duplicados, pero la inmensa mayoría no. Al desbordamiento cognitivo de cada motor se le suma el ruido informacional de recursos repetidos tantas veces como buscadores interrogados.



EJEMPLO

Alguno de los metabuscadores más conocidos son:

- [All 4one](#): Interroga los motores de Altavista, Lycos, HotBot i Excite.
- [ZOO](#): Realiza les consulta simultánea de: Google, Yahoo i Truveo