



## TÉCNICAS DE BÚSQUEDA Y USO DE LA INFORMACIÓN

## TEMA 3

## La recuperación de información en entornos tecnológicos

Página | 1

## OBJETIVO

- Capacitar al alumno en la utilización de herramientas y estrategias de búsqueda para la recuperación de información significativa en entornos tecnológicos.

## 1) Formas básicas de obtener información: navegación e interrogación (navegar y buscar)

El proceso de **recuperación de información**, que adopta la forma de explotación de Bases de Datos en el **entorno web**, abarca el conjunto de tareas que el usuario de información realiza para localizar y acceder a las fuentes de información que necesita para satisfacer su necesidad informativa (resolver su problema de información). En dicho entorno Web, son dos las formas básicas de obtener información:

- Mediante **NAVEGACIÓN**: consiste en obtener información visualizando las informaciones que presenta una página web principal y que, organizadas de un modo concreto, suelen presentar enlaces (links) a otros contenidos. Suele ser útil cuando no se conoce bien un tema y para obtener “buenos” conjuntos de recursos de información especializados, cuando la página está realizada por documentalistas o profesionales de un ámbito concreto.
- Mediante **INTERROGACIÓN**: consiste en obtener información introduciendo en el formulario de búsqueda que ofrece una Base de Datos/página web una expresión de búsqueda, es decir, una o más palabras que traducen la necesidad de información. Ello permite obtener una lista de referencias sobre datos o documentos, dentro de los cuales están, en un orden concreto, algunas, varias o todas las palabras utilizadas en la búsqueda.

## 2) Navegar por la información: estructura de la información y algunas pautas útiles

La información navegable se puede estructurar de tres maneras:

- a) **De forma secuencial** mediante **listas** (organizadas por orden alfabético, numérico, temático, de importancia...). Ejemplo: <http://www.doaj.org/>
- b) **De forma jerárquica**, mediante **árboles** (se jerarquiza el orden del contenido principal - La estructura de un sitio web es igual que un árbol que tiene ramas, que sostienen otras ramas para finalmente llegar a las hojas). Ejemplo: <http://www.doaj.org/>
- c) **De forma asociativa**, mediante **mapas** (los mapas listan los contenidos (páginas) de un sitio web y, en ellos, los términos están relacionados entre sí aunque no tengan una dependencia jerárquica estricta). Ejemplos característicos en la web son los *mapas de sitio o mapas web* que, en numerosas ocasiones, se presentan en la portada del sitio web que se consulta (normalmente, en uno de los bordes de la página). Ejemplos:

Mapa del sitio de la UE:

- [https://europa.eu/european-union/abouteuropa/sitemap\\_es](https://europa.eu/european-union/abouteuropa/sitemap_es)

Mapa del sitio de la Secretaría General de Instituciones Penitenciarias:

- <http://www.institucionpenitenciaria.es/web/portal/mapa.html>

**Navegar por la información** a partir de una página principal en la que existe mucha información, enlaces y opciones a seguir, puede resultar difícil, pero algunas **pautas básicas** pueden ser de utilidad en el proceso de recuperación de información por navegación:

- ✓ Observar detenidamente la página web y, especialmente, sus bordes, donde suelen ubicarse sus distintos menús y submenús.
- ✓ Cuando se está buscando una información concreta, hay que considerar si existe alguna categoría genérica (enlace) que pueda dar acceso al contenido que se busca. Cuando este sea el caso, no basta con quedarse en dicha categoría, siendo conveniente revisar otros enlaces para comprobar si alguno de ellos puede dar también acceso a la información que se necesita (no es conveniente, nunca, quedarse con la primera información que se encuentra, por defecto).
- ✓ Observar los “accesos directos” que en muchos casos ofrecen los sitios web, ya que suelen conducir a la información considerada más importante, más útil o, simplemente, más consultada (este tipo de información, no obstante, puede aparecer también en el centro o en los extremos de la página).
- ✓ Cuando se buscan Organismos vinculados a un tema de interés o investigación, pero no se tiene claro o se desconocen los que existen, es recomendable acceder a un Organismo mayor del que puedan depender (grandes instituciones, como son los ministerios, suelen tener organismos dependientes y enlaces a las páginas principales de éstos, desde su página principal).
- ✓ No quedarse sólo en la primera pantalla que se ve. El tamaño de esa primera pantalla depende de muchos factores (incluido el ordenador que se usa). Es posible que la información esté al final, o que la pantalla que se ve, sea más ancha de lo que a primera vista parece.
- ✓ Considerar la información contenida en la dirección o URL de la página/sitio web, ya que, a veces, permite “navegar” a través de ella (en algunos casos, la URL incluye una lista de categorías y subcategorías previas a la páginas; la URL se puede ir recortando para ver qué se puede conseguir). Un ejemplo: [http://europa.eu/youreurope/citizens/consumers/financial-products-and-services/index\\_es.htm](http://europa.eu/youreurope/citizens/consumers/financial-products-and-services/index_es.htm)
- ✓ Intentar distinguir qué parte del contenido de la página realmente aporta la información que se busca (evitar el RUIDO: contenido que no aporta información al usuario). La página principal de numerosos organismos tienen buena parte de sus contenidos ocupados, especialmente en el centro de la página, con noticias, notas de prensa, fotos de sus responsables...

### 3) Buscar información por interrogación: la importancia de la representación del contenido de los recursos de información para la recuperación informativa

En la **recuperación de información por interrogación**, resulta fundamental conocer la forma en que los datos formales y de contenido de los recursos de información aparecen en los distintos **campos** de los registros de **cada base de datos** (ver TEMAS 1 y 2), ya que **es sobre los campos** (autor, título, materia...) **sobre los que se efectúa la búsqueda de información** (las informaciones contenidas en los campos sobre los que frecuentemente se suelen hacer búsquedas a través de los motores de búsqueda de las propias bases de datos, quedan recogidas en los distintos índices de ésta: autores, materias...).

La utilidad que, especialmente, ofrece **la búsqueda de información por materias**, resalta el papel que juegan los profesionales / expertos en la creación de las bases de datos, ya que, en cuanto responsables de

**determinar** (representar) **el contenido de los recursos de información**, facilitan la recuperación significativa de información.

La descripción del contenido temático de un recurso informativo se realiza por los profesionales a través de la llamada **INDIZACIÓN**, proceso que:

Página | 3

- Consiste en **extraer** (partiendo de elementos como el título, índice, resumen....) todas aquellas **nociones significativas o términos que representan el contenido** de un recurso informativo.
- Puede realizarse utilizando el **lenguaje natural** (lenguaje habitualmente empleado) o un **lenguaje documental** (lenguaje convenido profesionalmente).
- Da lugar a la identificación de **palabras clave / key words** (lenguaje natural) o **descriptores temáticos** (lenguaje documental).

Para realizar este proceso se utilizan, entre otros instrumentos (propios del ámbito documental), los siguientes:

- **Listas de encabezamientos de materia:** son listados controlados de materias, basados en el lenguaje natural, que de forma convenida permiten a las bibliotecas y servicios de información adscribir una materia a los documentos que albergan. Suelen utilizarse, junto a la CDU, para completar los datos relativos a los contenidos en un registro bibliográfico (ver tema 1).
- **Sistemas de clasificación:** propios del lenguaje documental, representan las materias mediante un símbolo numérico, alfanumérico o alfabético y agrupan los conceptos vinculados a ellas siguiendo un **orden jerárquico** definido por el estado de los conocimientos existente en el momento en que se elabora (los conceptos más específicos se sitúan dentro de los conceptos más generales, presentando una estructura en forma de árbol (categoría – subcategoría - subcategoría...)).

Pueden ser:

- **Enciclopédicos** (en España la **CDU: Clasificación Decimal Universal**) —en ella se basa la signatura topográfica que permite localizar los documentos en las unidades de información—

#### CDU. Clasificación Decimal Universal

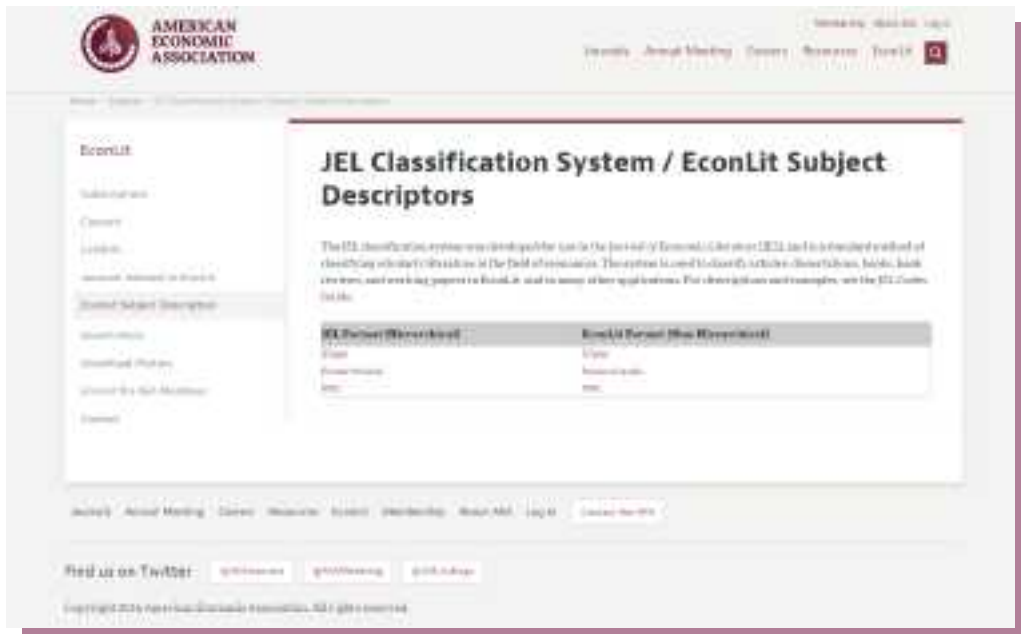
0	Generalidades. Ciencia y conocimiento. Organización. Información. Documentación
1	Filosofía. Psicología
2	Religión. Teología
3	Ciencias sociales. Estadística. Política. Economía. Comercio. Derecho
4	(Vacante)
5	Matemáticas. Ciencias Naturales
6	Ciencias aplicadas. Medicina. Tecnología
7	Bellas artes. Espectáculos. Deportes
8	Lenguaje. Lingüística. Literatura
9	Geografía. Biografías. Historia

**Periodismo: 070 (CDU)**

**Comunicación Audiovisual: 316.774 (CDU)**

- **Especializados** (Ejemplo: en el área de Economía el más utilizado es el **JEL: Journal of Economic Literature Classification System**)

**JEL Codes Guide:** <https://www.aeaweb.org/econlit/jelCodes.php>



The screenshot shows the American Economic Association website. The main content area is titled "JEL Classification System / EconLit Subject Descriptors". It includes a brief description of the JEL system and a table comparing JEL codes with EconLit subject descriptors.

JEL Subject (EconLit Code)	EconLit Subject (JEL Classification)
Microeconomics	J1-J4
Macroeconomics	J5-J8
International Trade and Finance	J9

At the bottom, there is a section for "Find us on Twitter" with links to @AEAonline, @AEAmeeting, and @AEApublic. The footer contains copyright information for 2015.

- **Tesauros:** constituyen la principal herramienta para el análisis, descripción y recuperación de información en el entorno tecnológico, realizadas en lenguaje documental por los profesionales de la información. Son vocabularios de términos controlados (**descriptores**) que se aplican a un dominio concreto (área de conocimiento o disciplina) y muestran las relaciones jerárquicas, de equivalencia y asociativas que se dan entre ellos (las relaciones sirven para delimitar el contenido semántico de los descriptores y para expresar los vínculos que les relacionan entre sí):
  - **Relaciones Jerárquicas.** Se indican mediante **TG** (Término Genérico) y **TE** (Término Específico). Expresan los distintos niveles, género/especie o todo/parte, que se dan entre los descriptores, lo que permite situarlos en un contexto preciso. La relación se establece de forma vertical entre todos los descriptores de una misma clase, organizados de más genéricos a más específicos.
  - **Relaciones de equivalencia.** Se indican mediante **USE** y **UP** (Usado por). Reflejan la relación entre términos aceptados (descriptores) y no aceptados (no descriptores). Cuando delante de un término aparece USE, el término que sigue a esta etiqueta es el descriptor que se debe usar. Cuando lo hace UP, el término que sigue a esta etiqueta es el no descriptor.
  - **Relaciones asociativas.** Se indican mediante **TR** (Término Relacionado). Ponen en relación descriptores que no tienen jerarquía entre ellos.

Cuando un descriptor ofrece dudas en cuanto a su significado o aplicación se utiliza **NA (Nota de Alcance)** para explicar su definición (a fin de deshacer ambigüedades) o para delimitar cuándo debe utilizarse y cuándo no.

Ejemplo: Tesauro de Economía

<b>Macroeconomía</b>	
UP	<a href="#">Análisis macroeconómico</a>
UP	<a href="#">Teoría macroeconómica</a>
TG	<a href="#">Teoría económica</a>
TE	<a href="#">Ciclos económicos</a>
TE	<a href="#">Empleo agregado</a>
TE	<a href="#">Inflación</a>
TE	<a href="#">Magnitudes macroeconómicas</a>
TE	<a href="#">Modelos macroeconómicos</a>
TE	<a href="#">Política macroeconómica</a>
TR	<a href="#">Consumo</a>
TR	<a href="#">Demanda</a>
TR	<a href="#">Economía del bienestar</a>
TR	<a href="#">Equilibrio económico</a>
TR	<a href="#">Inflación</a>
TR	<a href="#">Keynesianismo</a>
TR	<a href="#">PIB</a>
TR	<a href="#">Producción</a>
TR	<a href="#">Producto nacional bruto</a>
TR	<a href="#">Producto nacional neto</a>
TR	<a href="#">Renta nacional</a>
TR	<a href="#">Tablas input-output</a>
TR	<a href="#">Teoría del crecimiento</a>
TR	<a href="#">Utilidad</a>
EN	Macroeconomics
FR	Macroéconomie

<b>Madrid Interbank Offered Rate</b>	
USE	<a href="#">MIBOR</a>

<b>Magnitudes monetarias</b>	
NA	Activos líquidos de toda clase de empresas, personas jurídicas y naturales

#### 4) La recuperación de información por interrogación como proceso<sup>1</sup>

Una vez identificada la necesidad de información, el proceso de recuperación informativa implica habitualmente los siguientes pasos:

1. Concretar el **tema** (definir el objetivo, alcance y limitaciones del tema sobre el que se desea obtener información)
2. Seleccionar los términos que identifiquen mejor el objetivo de la búsqueda de información (**palabras clave / descriptores**). Con ayuda de tesauros, diccionarios..., obtener también términos relacionados que permitan disponer de términos alternativos por los que poder ampliar la búsqueda, sinónimos, etc. y la traducción de los términos a otros idiomas.
3. Seleccionar las **bases de datos** concretas a las que se desea interrogar.

Tener en cuenta que:

- o La selección de las bases de datos a las que interrogar debe determinarse conociendo el **TIPO DE BÚSQUEDA QUE SE NECESITA REALIZAR** y las **PRINCIPALES FUENTES DE INFORMACIÓN** a las que se puede acudir para satisfacer una necesidad de información (**aproximación a un tema: buscadores generales, obras de referencia...**; **obtención de datos específicos: estadísticas...**; **identificación, localización y obtención de documentos: catálogos de bibliotecas, repositorios institucionales...**; **conocer con cierta profundidad la información que hay sobre un tema hasta la fecha: revistas especializadas...**)

<sup>1</sup> Vincular apartado al Tema 1 y tener presente para el Trabajo Final

- Las bases de datos de los **buscadores generales**, como **Google** o **Yahoo** (ver Tema 2), son útiles para búsquedas genéricas o exploratorias por las que aproximarse a un tema pero **no son suficientes**.
4. Plantear una **estrategia de búsqueda** (la forma de combinar los **términos clave** identificados **utilizando operadores de búsqueda y/o los campos específicos de una interfaz** de búsqueda).

Tener en cuenta que:

- Las bases de datos ofrecen la posibilidad de buscar en lenguaje natural o en **lenguaje controlado**. En este último caso, se tendrán que **traducir los términos elegidos del lenguaje natural al lenguaje documental utilizado por la base de datos**, por lo que es conveniente seguir la información específica de uso que aportan (**Ayudas**) --a menudo incluyen *tesauros* que facilitan la traducción al lenguaje documental y permiten ampliar los términos de búsqueda--.
  - La utilización de **los motores de búsqueda de las bases de datos permiten afinar mejor las búsquedas que los directorios o mapas de navegación** que permiten explorar a través de enlaces, agrupaciones de recursos previamente establecidas (ver epígrafe 1 y 2 de este tema).
5. Ejecutar la estrategia, obtener y **evaluar los resultados** (mediante las llamadas **medidas de recuperación**) y, en su caso, replantear la estrategia planteada.
6. Seleccionar los resultados y obtener y **consultar los documentos** originales.

#### Observaciones:

- La **necesidad informativa** de un usuario **no se puede traducir con exactitud en una estrategia** de búsqueda.
- Las **respuestas** que ofrecen las herramientas de búsqueda a las estrategias de búsqueda formuladas no presuponen fiabilidad y rigor, **sólo responden a la estrategia planteada**; la **valoración del usuario** es fundamental
- Los **resultados** obtenidos en una búsqueda **necesitan complementarse con el acceso al documento original, su revisión y su selección, o rechazo**, en función de **criterios establecidos por el usuario**

## 5) Operadores de búsqueda

Son **vocablos o símbolos que permiten realizar las combinaciones entre los términos** elegidos para la **búsqueda** de información. **Cada combinación de términos, constituye una estrategia de búsqueda** y ésta variará en función del tipo de información que se necesite y el sistema de información (base de datos) sobre el que se aplique.

### Tipos

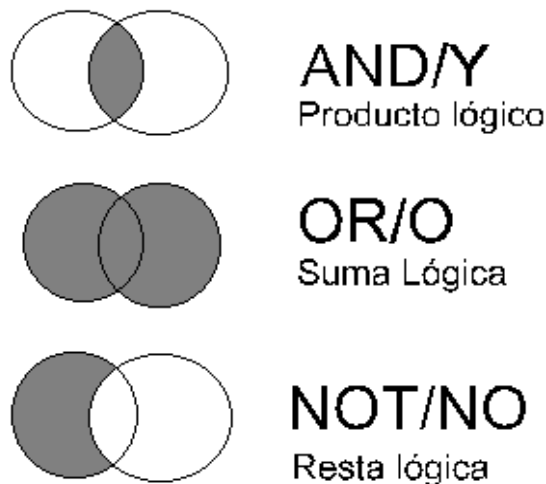
#### OPERADORES LÓGICOS

- Se denominan booleanos en honor a George Boole, precursor del Álgebra de conjuntos.
- El principio que rige su utilización es que las relaciones entre conceptos pueden expresarse como relaciones matemáticas entre conjuntos:

**AND ( Y )**: Operador Producto / Intersección. Cuando se utiliza este operador entre dos términos de búsqueda, se pide a la base de datos que recupere **documentos que contengan ambos términos**.

**OR ( O )**: Operador Suma. Cuando se utiliza este operador entre dos términos de búsqueda, se pide a la base de datos que recupere **documentos que contengan cualquiera de los dos términos y ambos términos**.

**NOT ( NO )**: Operador Resta. Cuando se utiliza este operador entre dos términos de búsqueda, se pide a la base de datos que recupere **documentos que contengan uno de los términos y no el otro**.



Estos operadores, a su vez, pueden combinarse entre sí para generar operaciones más complejas que se ejecutan siguiendo la lógica matemática.

#### LIMITACIONES

- Siempre se plantean en términos de absoluto (presente/ausente), sin considerar el peso específico del término en el contexto.
- Exigen un alto valor de precisión en los términos utilizados
- Requieren claridad en la composición de las expresiones a buscar.

#### OPERADORES DE PROXIMIDAD

- Permiten buscar por dos términos, separados por un número determinado de palabras. Consideran que la cercanía entre términos puede reflejar entre ellos una íntima relación:

**NEAR (CERCA)**: es un operador un poco parecido al operador booleano AND, pero a diferencia de éste, NEAR exige que entre dos términos no existan más de 10 palabras:

- NEAR1 (Una palabra)
- NEAR2 (Dos palabras)
- NEAR3 (Tres palabras)

**ADJ (SEGUIDO)**: es un operador un poco parecido al operador booleano AND. Busca dos términos consecutivos (adyacentes, juntos).



## OPERADORES DE COMPARACIÓN

- Especifican el rango de búsqueda, fijando unos límites para la misma, que pueden ser tanto numéricos como alfabéticos:

**> mayor**

**< menor**

**= Igual**

**>= mayor o igual**

**<= menor o igual**

Se utilizan principalmente para obtener datos numéricos y para recuperar documentos en las bases de datos bibliográficas según la fecha de publicación.

## OPERADORES DE TRUNCAMIENTO

- Se les denomina también “**máscaras**”
- Se utilizan en situaciones en las que es necesario utilizar, no un término simple, sino también, sus derivados, fijados por prefijación o sufijación, mínimas variantes léxicas...
- Se representan mediante símbolos: \*, \$, ?
- Su presencia puede sustituir a un carácter o a un conjunto de caracteres, situados a la izquierda, dentro o a la derecha del término en cuestión.

## OPERADOR DE EXACTITUD: Comillas (“ .... ”)

- Suele constituir una de las opciones más comunes de búsqueda en los llamados buscadores generales (Google, Yahoo...).
- Exige que la secuencia de palabras sea idéntica en el resultado de la búsqueda (presupone que la frase debe aparecer en el texto recuperado).
- Se aconseja utilizar frases cortas que contengan términos que aparezcan con frecuencia en los documentos consultados.

## 4) Medidas de recuperación (Valoración de los resultados de las búsquedas)

Las medidas de recuperación permiten evaluar el grado de éxito del proceso de recuperación de información. Las más importantes son:

### ► RELEVANCIA

- Documentos relevantes son los documentos que, potencialmente, pueden responder a una estrategia de búsqueda determinada. La relevancia mide el grado de adecuación de los documentos recuperados a una determinada estrategia.

► **PRECISIÓN** (¿cuánto de lo que se ha recuperado es relevante?)

- Tasa que mide la relación entre el número de documentos relevantes recuperados y el número total de documentos recuperados (% de documentos relevantes recuperados del total de documentos recuperados).
  - **Cálculo de la Tasa de Precisión:**  $P = (A/A + B) \times 100$ , siendo:
    - A = Documentos recuperados relevantes
    - B = Documentos recuperados no relevantes
    - A+B = Total de documentos recuperados

► **RUIDO** (¿cuánto de lo que se ha recuperado no es relevante?)

- Tasa que mide la relación entre el número de documentos no relevantes recuperados y el número total de documentos recuperados (% de documentos no relevantes recuperados del total de documentos recuperados).

► **SILENCIO** (¿cuántos documentos relevantes han quedado por recuperar?)

- Tasa que mide la relación entre el número de documentos relevantes extraídos y el número de documentos relevantes existentes. (% de documentos relevantes no recuperados del total de documentos relevantes). Es difícil de calcular, por la dificultad práctica de conocer el número total de documentos relevantes existentes en un sistema para una necesidad de información dada.

### Recomendaciones generales

- Si una búsqueda ofrece **muchos resultados**:
  - Buscar mejores términos de búsqueda
  - Utilizar operadores de proximidad y eliminar truncados
  - Utilizar descriptores
  - Buscar en campos específicos ...
- Si una búsqueda ofrece **pocos resultados**:
  - Ampliar búsquedas en otros campos
  - Eliminar los conceptos menos relevantes
  - Utilizar truncados y eliminar operadores de proximidad
  - Ampliar la búsqueda a cualquier tipo de documento ...