

The seal of the University of Valencia is a circular emblem. It features a central shield with a crown on top. The shield is divided into four quadrants: the top-left and bottom-right are white, the top-right and bottom-left are black. The shield is flanked by two smaller shields, each with a crown on top. The left shield has a white background with a black cross, and the right shield has a black background with a white cross. The entire seal is surrounded by a circular border containing the Latin text "GRA REX ARAGONVM • ALEXANDER PRIMI VALENTINVS DEI FIDELIUM".

Boletín de prácticas
Práctica 2: Búsquedas en Internet

INFORMÁTICA 1

Curso 2004-2005

Biblioteconomía y Documentación

1.- INTRODUCCIÓN

En esta práctica se van a desarrollar las aptitudes básicas de búsqueda de información en Internet. A lo largo de la práctica el alumno aprenderá técnicas para localizar la información deseada y refinar los resultados obtenidos tanto en lo referente a eliminar los resultados no pertinentes como a obtener resultados pertinentes adicionales que no se habían logrado.

2.- OBJETIVOS DE LA PRÁCTICA

Se pretende que el alumno, mediante la realización de esta práctica, desarrolle las siguientes habilidades:

- Definir de forma precisa el objeto de la búsqueda a partir de la información de partida y sus propios conocimientos sobre la materia.
- Elegir las herramientas de consulta adecuadas en función de las características de la información a localizar, evitando la costumbre habitual de usar siempre el mismo buscador.
- Elegir las mejores estrategias de consulta.
- Aprender a evaluar y seleccionar la información encontrada según criterios de pertinencia, actualidad y calidad.
- Refinar las estrategias de consulta según los resultados obtenidos.

3.- HERRAMIENTAS DE BÚSQUEDA

Buscar información en Internet no es sólo buscar en la World Wide Web. Existe mucha información útil y accesible a través de la red que no está en páginas web. Sin embargo, dado el crecimiento experimentado por la web, el volumen de información contenido en páginas web o referenciado por alguna de ellas es cada vez mayor.

Existen una serie de problemas cuando queremos localizar información en la web:

- El volumen de información es muy grande. Mantener referenciadas todas las páginas web es casi imposible. No existe ninguna herramienta que mantenga referencias a toda la web, por lo que ninguna herramienta nos podrá asegurar una eficiencia total en la búsqueda.
- La información en Internet es muy volátil. Las páginas cambian, aparecen o desaparecen continuamente. Aunque la mayoría de las herramientas de consulta actualizan permanentemente sus bases de datos, a menudo contienen referencias obsoletas, por lo que mucha información indexada ya no existe o ha cambiado de ubicación.
- Existe mucha información “basura”. Dado que no existe apenas control acerca de la información a publicar, cualquiera puede publicar prácticamente cualquier cosa, por lo que a la dificultad de localizar las páginas pertinentes se añade la dificultad de distinguir la información de calidad.

Con todo, a pesar de estos problemas existen multitud de herramientas que nos permiten mejorar nuestra capacidad de localización de información relevante en la red. A continuación veremos los tipos principales de herramientas de consulta existentes en la actualidad: motores de búsqueda, metabuscadores y directorios temáticos.



3.1 Motores y metabuscadores

Los motores son bases de datos que guardan información de páginas web seleccionadas. Normalmente, estas páginas web son recopiladas mediante programas llamados *arañas*, que recorren la web de forma automática siguiendo los enlaces de las páginas visitadas y guardando información acerca de las mismas así como de sus enlaces. Los motores permiten búsquedas de información mediante palabras clave.

Algunos ejemplos de motores de búsqueda pueden ser:

- Altavista: <http://www.altavista.com>
- Google: <http://www.google.com>

Los metabuscadores son interfaces unificadas que acceden a la vez a las bases de datos de varios motores. Los metabuscadores no contienen realmente referencias a páginas, sino solamente información de acceso a los motores a los que consultan. El proceso que siguen es, por tanto, recoger la consulta del usuario, lanzarla sobre una serie de motores, obtener las listas de enlaces de resultado de cada uno de ellos y reorganizar dichas listas para mostrárselas al usuario de forma homogénea.

Algunos ejemplos de metabuscadores de búsqueda son:

- Metacrawler: <http://www.metacrawler.com>
- Biwe: <http://multibuscador.biwe.es>
- Metaíndice: <http://metaindice.com>

3.2 Directorios temáticos

Los directorios temáticos son bases de datos de sitios web estructurados según categorías con estructuras jerárquicas yendo de lo más general a lo más especializado. Para cada categoría existen subcategorías más específicas así como enlaces a los sitios web. Normalmente los sitios web son recopilados, analizados y clasificados por expertos. Es interesante observar que se recopilan sitios web y no páginas aisladas, por lo que habitualmente encontraremos la página inicial de acceso a un sitio y no cada una de las subpáginas existentes en el mismo.

Las búsquedas en estos directorios se realizan tanto navegando por las categorías temáticas como buscando términos que aparezcan en éstas o en las descripciones de los recursos indexados. Normalmente no permiten búsquedas de contenido; tan sólo se puede buscar en la información recopilada, en datos tales como URL, título, campos meta, descripción del indexador, categorías por materias, etc.

Entre los directorios más conocidos podemos destacar:

- Yahoo <http://www.yahoo.com>
- Yahoo español <http://www.yahoo.es>
- Terra <http://www.terra.es>
- About: <http://about.com>

4.- ESTRATEGIAS DE BÚSQUEDA

Para obtener la información que necesitamos en Internet deberemos elegir la herramienta o herramientas de consulta adecuadas y, una vez en ella, desarrollar una



estrategia de consulta adecuada. La información se obtendrá mediante el uso de estrategias de búsqueda heurísticas y difícilmente automatizables. Por ello, se pueden establecer algunas directrices a seguir en el desarrollo de las búsquedas, pero es muy difícil establecer un plan seguro para obtener la máxima eficacia en las consultas.

4.1 Fases del desarrollo de una búsqueda

Si bien es casi imposible dar un guión que sirva para obtener la máxima eficacia en las consultas, sí se pueden plantear una serie de fases a seguir en el desarrollo de las mismas para maximizar en lo posible la calidad y el volumen de los resultados:

1. Análisis de la necesidad de información: Aunque pueda resultar obvio, debemos preguntarnos qué es exactamente lo que buscamos, qué profundidad buscamos en las respuestas, qué datos de partida tenemos y qué conocimientos personales sobre la materia tenemos.
2. Elección de un punto de partida: Debemos elegir en primer lugar la herramienta de consulta adecuada. Si la herramienta elegida es un motor o meta-motor, deberemos elegir los términos a buscar (sin olvidarnos de sus posibles sinónimos). Si la herramienta elegida fuese un directorio temático deberemos elegir las categorías temáticas en que creemos puede estar englobada la información que buscamos.
3. Adaptación de la estrategia de consulta en función de los resultados obtenidos: A medida que vayamos obteniendo resultados, podremos ir modificando nuestra estrategia, bien por la navegación a través de los enlaces obtenidos, bien por el uso de términos alternativos que nos hayan aparecido en los resultados parciales.
4. Evaluación de los resultados obtenidos: Puesto que no existen apenas restricciones para la publicación de información en Internet, deberemos evaluar desde un punto de vista crítico los resultados obtenidos atendiendo a criterios de autenticidad, pertinencia, actualidad y veracidad.

Evidentemente, repetiremos las fases tres y cuatro hasta llegar a unos resultados que nos parezcan óptimos. La fase cuatro es crítica en cuanto al hecho de que puede publicarse información incompleta, parcial o falsa, deliberadamente o no. Por ello, deberemos recurrir a técnicas que nos permitan vislumbrar de alguna forma la fiabilidad de los datos obtenidos, analizando características como:

- Análisis de la URL: Dependiendo de la forma de la URL podemos conjeturar si se trata una página personal o institucional. Por ejemplo, podemos ver el dominio del servidor para averiguar la ubicación de la página.
- Verificación de la autoría de la página: Podemos realizar una búsqueda paralela del autor o autores de la página para saber más.
- Evaluar qué páginas tienen enlaces a la que estamos evaluando. Si una página que consideramos altamente fiable tiene enlaces a la página a evaluar podemos *fiarnos* algo más de ésta última.
- Analizar las motivaciones políticas, personales, sociales o económicas de la publicación de la página ¿qué buscaba el autor al publicar la página?



4.2 Sentencias de búsqueda: operadores básicos.

La sintaxis y forma de las sentencias de búsqueda dependerán de la herramienta que hayamos elegido. En cualquier caso, antes de realizar nuestras consultas deberemos tener presentes las funcionalidades del recurso que estemos utilizando, ya que no todos los buscadores permiten usar todos los operadores de consulta habituales y desde luego la sintaxis de los mismos no está estandarizada. Por ejemplo, uno de los operadores más habituales es el truncamiento, pero sin embargo Google, uno de los buscadores más utilizados en la actualidad, no lo soporta.

A continuación revisaremos las sentencias y operadores más habituales:

4.2.1 Sentencia simple

La sentencia simple consiste en indicar un único término a buscar, como por ejemplo, *arma*. Esta sentencia nos devolvería una lista de las URL's que contengan en alguna de las partes indexadas dicho término. No debemos olvidar qué contenidos indexa el buscador que estemos usando. El uso de términos simples nos generará grandes volúmenes de resultados, si bien la cantidad de resultados no pertinentes será alta en la mayoría de los casos.

4.2.2 Sentencia múltiple

La sentencia múltiple consiste en indicar más de un término simple, como por ejemplo *armas destrucción masiva*. Esta sentencia nos devolvería una lista de las URL's que contengan en alguna de las partes indexadas los términos referidos en cualquier orden y posición. Esto supone que nos encontraría tanto páginas que contengan la frase “armas de destrucción masiva” como “destrucción masiva de armas”, a pesar de tener significados bien diferentes. Incluso encontraríamos páginas que contuviesen los tres términos en párrafos diferentes y en cualquier orden.

4.2.3 Búsqueda por frase

Si deseamos buscar, como en el ejemplo anterior, una frase concreta y no en general los términos que contiene, deberemos utilizar la búsqueda por frase. En multitud de buscadores la sintaxis para indicar este tipo de búsqueda es poner la frase deseada entre comillas, aunque algunas herramientas contienen una caja de texto específica para esta función. Por ejemplo, en Google la consulta “*armas de destrucción masiva*” nos localizará los recursos que contengan exactamente esta frase. Obsérvese que se busca la frase exacta, incluyendo los términos vacíos que haya, como en este caso la partícula “de”.

4.2.4 Truncamiento

Si estuviésemos buscando información sobre un determinado término y obtenemos pocos resultados, puede ser interesante ampliar la búsqueda a otros términos con la misma raíz o, en general, a términos que cumplan un determinado patrón. De esta forma, podríamos extender nuestra búsqueda de *armas* a los términos *arma*, *armamento*, *armamentístico*, *armado*, etc. Los operadores más habituales de truncamiento suelen ser *,\$,#. Hemos de tener en cuenta que muchos buscadores no soportan truncamientos en absoluto y que otros lo soportan con restricciones. Además, la función de los operadores de truncamiento si existen varía de unos a otros.



4.2.5 Operadores booleanos

Los operadores booleanos permiten combinar los conjuntos de resultados obtenidos de las búsquedas parciales mediante aritmética de conjuntos. De este modo, podremos usar habitualmente:

- AND: Equivalente a la intersección de conjuntos. La búsqueda *armas AND destrucción* devolverá los recursos que contengan ambos términos, es decir, realizará la intersección del conjunto de páginas que contengan el primer término con el conjunto de páginas que contengan el segundo.
- OR: Equivalente a la unión de conjuntos. La búsqueda *armas OR destrucción* devolverá los recursos que contengan alguno de los dos términos, es decir, realizará la unión o suma del conjunto de páginas que contengan el primer término con el conjunto de páginas que contengan el segundo. Es importante recordar que la unión no es exclusiva, a pesar de que en lenguaje natural pueda implicarlo a menudo, sino una suma simple de ambos conjuntos.
- NOT: Equivalente a la resta de conjuntos. La búsqueda *armas NOT destrucción* devolverá los recursos que contengan el primer término pero no el segundo, es decir, devuelve el contenido del conjunto de páginas que contengan el primer término exceptuando aquellas que pertenezcan al conjunto de páginas que contengan el segundo.

En el uso de los operadores booleanos es importante recordar las precedencias de los operadores y el uso de los paréntesis para forzar precedencias diferentes.

Dependiendo del buscador usado, podremos usar o no estos operadores directamente en las búsquedas o en cajas de texto específicas.

4.2.6 Inclusión y exclusión de términos

Algunos motores ofrecen una variación de los operadores booleanos AND y NOT para incluir o excluir términos específicos. Por ejemplo, en Google, si precedemos un término o frase del signo "+", requerirá que ese término o frase está presente en todos los recursos localizados. Si por el contrario usamos el signo "-", garantizaremos que dicho término no está presente en ninguno de los recursos localizados. Por ejemplo, la sentencia *+gas +VX -butano* nos localizará los recursos que contengan los dos primeros términos pero no el tercero.



5.- DESARROLLO DE LA PRÁCTICA

La práctica constará de dos fases: en la primera, el alumno deberá utilizar motores de búsqueda o metamotores para localizar información específica que permita satisfacer necesidades de información por parte de usuarios imaginarios. En la segunda parte, el alumno deberá utilizar dos directorios concretos para localizar sitios que traten sobre temas dados, comparando los resultados y su pertinencia según las dos herramientas utilizadas.

5.1 Fase 1: Búsqueda de información mediante motores de búsqueda o metamotores:

El alumno editará un documento en Microsoft Word en el que incluirá los resultados de las consultas realizadas en forma de fichas. Previamente al desarrollo de las búsquedas el alumno creará una ficha tipo con al menos las siguientes partes:

Pregunta:

Estrategia de consulta usada (para cada paso)

Buscador usado y su tipo

Términos usados

Términos desechados

Razonamiento usado en este paso

Páginas visitadas

Resultados seleccionados como óptimos

Respuesta

URL's de los documentos con la respuesta

Evaluación de la calidad de la respuesta obtenida

Para cada una de las preguntas que se enumeran más adelante, el alumno deberá llenar una ficha en dicho documento con los datos anteriormente citados. Se deberán describir los pasos seguidos para la consulta razonando cada una de las decisiones tomadas respecto de la estrategia de búsqueda, así como evaluar la pertinencia y calidad de los resultados finalmente elegidos. La elección del buscador o buscadores a utilizar es libre por parte del alumno.

Las preguntas a contestar son las siguientes

1. ¿Cuál es el país menos corrupto del mundo según el Índice de Percepciones de Corrupción (IPC) en 2003? ¿Qué índice tiene España en 2003?
2. ¿Qué es el *garum*?
3. Localiza una lista de los elementos declarados Patrimonio de la Humanidad en España por la organización de las Naciones Unidas dedicada a la educación, la ciencia y la cultura.
4. ¿Qué son los premios Stella? Localiza algún sitio web donde aparezcan los ganadores de los últimos años.
5. ¿Qué dice la famosa frase del presidente de Estados Unidos entre 1933 y 1945 acerca de cierto dictador nicaragüense?



6. Localiza dos ediciones íntegras del Quijote accesibles vía web.
7. ¿Dónde nació el autor de novelas de ciencia ficción George H. White?
8. ¿Quién fundó Google? ¿Cuántas páginas web indexa Google actualmente?
9. Localiza un libro en español que trate sobre un lenguaje de programación que debe su nombre al nombre de pila de la hija de un famoso poeta británico nacido en 1788 y muerto en 1824.
10. Localiza tres traducciones a tres idiomas diferentes del más conocido discurso de cierto activista estadounidense de los derechos civiles asesinado en Memphis en 1968.
11. ¿Qué antivirus es el mejor contra los virus tipo *hoax*?
12. ¿Qué estudios de postgrado en Biblioteconomía y Documentación existen actualmente en España?
13. ¿En qué año nació el dirigente que ordenó por primera vez el uso de gas mostaza en el territorio del actual Irak? ¿Qué importante premio obtuvo este dirigente en 1953?
14. ¿Quién inventó la heroína? ¿Cuál era su uso original?
15. ¿Sobre qué sistemas operativos funciona el software Greenstone? ¿Para qué sirve este software?

5.2 Fase 2: Búsqueda de información mediante directorios:

En esta segunda fase analizaremos dos directorios realizando tres consultas idénticas y comparando los resultados obtenidos para cada uno. Para ello, elegiremos tres temas de una lista dada y realizaremos las búsquedas sobre los **directorios** Yahoo! y Google.

Podemos elegir un tema de cada uno de estos tres bloques:

- Biblioteconomía:
 - o Evaluación de sistemas de información.
 - o Evaluación de motores de búsqueda.
 - o Sistemas de gestión de bases de datos documentales.
 - o Servicios de bibliotecas universitarias a través de Internet
- Informática:
 - o Recursos para el aprendizaje de HTML.
 - o Sistemas de automatización de bibliotecas.
 - o Accesibilidad de páginas web.
 - o Aplicaciones de gestión de referencias bibliográficas.
- Varios:
 - o Cáncer de mama.
 - o Anorexia nerviosa
 - o Descarga de libros electrónicos
 - o Cómic español.



- Periódicos en España.
- Turismo rural.

Una vez elegidos los temas realizaremos las mismas consultas en cada uno de los buscadores, y evaluaremos los resultados obtenidos, obteniendo al menos los siguientes datos:

- Número de resultados obtenidos.
- Para los diez primeros resultados:
 - Porcentaje de enlaces rotos entre los recursos obtenidos.
 - Puntuaremos y sumaremos la relevancia de los recursos obtenidos de la siguiente forma:
 - (-1): Enlace roto
 - (0): Nada relevante
 - (1): Deficiente
 - (2): Bueno
 - (3): Muy bueno
 - (4): Excelente

Con los resultados obtenidos realizaremos una comparativa de ambos directorios, en la que también incluiremos información relativa a sitios de máxima relevancia conocidos pero no encontrados, adecuación de la relevancia sugerida a nuestras necesidades, etc. También deberemos consignar las coincidencias de resultados entre los diferentes buscadores.

Como en la primera fase, se deberá completar el estudio con un cuadro comparativo en el que se resuman los resultados obtenidos.

6.- NORMAS DE LA PRÁCTICA

- Cada equipo deberá enviar por correo electrónico **desde la dirección de correo de la Universidad de Valencia de uno de los miembros** el documento Word con los resultados de la práctica al profesor en un plazo máximo de una semana a partir de la segunda sesión de esta práctica. El profesor contestará en las cuarenta y ocho horas siguientes
- Si existiese algún problema referente al uso del correo electrónico deberá ser comunicado al profesor durante la sesión de prácticas.
- Los equipos constarán de un máximo de dos componentes. Los equipos que se formen para esta práctica deberán mantenerse para la práctica 3.
- En el documento Word con la memoria deberán constar el nombre, grupo de prácticas y DNI de los componentes del equipo.
- No se admitirán memorias entregadas por otro medio que no sea el correo electrónico..