

## Tema 2. La búsqueda documental

- 1.- Definición
- 2.-Tipos de búsqueda
- 3.-Lenguajes de interrogación
- 4.-Etapas de la búsqueda documental
  - 4.1.-Análisis de las necesidades de información: la entrevista de referencia
  - 4.2.-Metodología para establecer una estrategia de búsqueda

### 1.- Definición

UNE 50/113 Recuperación de información/Information retrieval/recherche de l'information. Acción, métodos y procedimientos para recuperar información específica previamente almacenada.

Proceso mediante el que se obtiene la información, los datos o las referencias que se ajustan a las especificaciones de una solicitud determinada realizada por un usuario.

### 2- Tipos de búsqueda

Tipos de búsqueda:

- En función del tiempo: Retrospectiva o prospectiva
- En función del soporte: manual o automatizada (online y offline)
- En función de la persona que la efectúa: personal, delegada
- En función del estado de conocimiento del usuario o distinto tipo de necesidad: ítem conocido, búsqueda temática (este punto, mucho más interesante, es de otros autores: Salvador y Angós)

#### En función del tiempo:

Búsqueda retrospectiva: Operación documental que permite la recuperación retrospectiva de un conjunto de documentos, de referencias o de datos relativos a un tema concreto. Se llama retrospectiva porque proporciona información relativa a publicaciones o acontecimientos sucedidos en un número determinado de meses o años anteriores a la fecha en la que se efectúa.

Búsqueda prospectiva: Operación documental que permite la obtención de un conjunto de documentos o de datos, o de información sobre acontecimientos relativos a un tema concreto a medida que la información va apareciendo en el tiempo. Las búsquedas prospectivas constituyen una operación de seguimiento de un tema. Se realizan mediante los servicios conocidos como de alerta. DSI

#### En función del soporte:

Búsqueda manual: recuperación de la información a partir de fuentes impresas.

Búsqueda automatizada: recuperación de la información a partir de soportes electromagnéticos, especialmente las bases de datos.

Búsqueda automatizada online. Acceso a la información a distancia y en tiempo real que permite la comunicación instantánea y conversacional con la información que contiene la base de datos. Vías de comunicación: teledocumentación, internet.

Búsqueda automatizada offline: Acceso a la información contenida en bases de datos de soporte óptico (cd-rom). Se pueden consultar mediante uso directo, mediante acceso vía internet o una red de bases de datos en cd-rom.

#### En función de la persona que la efectúa:

Búsqueda personal: la realiza el propio usuario.

Búsqueda delegada: la realiza el documentalista que actúa como intermediario entre el usuario y el sistema de información.

#### En función del estado de conocimiento del usuario o distinto tipo de necesidad:

Búsqueda de un ítem (unidad documental) conocido: El usuario necesita conseguir una referencia o un documento concreto del que conoce el autor o el título. Este tipo de búsquedas son muy sencillas de realizar y se pueden resolver consultando catálogos y bases de datos bibliográficas.

Búsqueda temática: El usuario necesita localizar material sobre un determinado tema o información que ayude a contestar a una pregunta. Puede subdividirse en tres categorías:

- a- Búsqueda de información que ayude a resolver una determinada pregunta que ayude a la toma de decisiones. En este caso es suficiente con proporcionar el dato factual o un único documento que contenga la información que necesita el usuario. *Concrete information need*
- b- Búsqueda de literatura publicada o del cuerpo de conocimiento existente sobre un determinado tema. Se resuelve con una búsqueda retrospectiva que puede ser más o menos exhaustiva en función de los objetivos perseguidos por el usuario. Son las búsquedas más frecuentes. *Problem oriented information need*
- c- Búsqueda de conocimiento actualizado: el usuario necesita estar al día de los nuevos avances y desarrollos que se producen en un campo, especialidad o tema determinados. Este tipo de búsqueda suelen ser posteriores a una búsqueda retrospectiva en la base de datos apropiada en la que el perfil de búsqueda utilizado ha dado unos resultados altamente relevantes. Este perfil se graba en la base de datos con un comando determinado que indica al sistema que envíe periódicamente los nuevos registros introducidos que coinciden con el perfil de búsqueda. *DSI*

### 3.- Los lenguajes de interrogación:

Las bases de datos documentales proporcionan el soporte necesario para el almacenamiento y la recuperación posterior de la información que contienen.

Al conjunto formado por la base de datos y el soporte necesario para almacenar y recuperar información se le denomina SGBD (Sistema de Gestión de Bases de Datos).<sup>1</sup>

Para recuperar la información almacenada, los sgdb, disponen de un lenguaje de interrogación que se compone de operadores que permiten elaborar consultas más o menos complejas. Los operadores más conocidos son los booleanos pero hay más tipos.

Para entender el funcionamiento de operadores es imprescindible conocer la estructura de una base de datos. Las bases de datos documentales se componen de registros, que representan al documento tanto físicamente (descripción física: título, autor, páginas, etc) como su contenido. Un registro es la representación abreviada de un documento.

Cada registro se divide en campos (unidad básica del registro). Un campo puede estar subdividido en subcampos mediante delimitadores de párrafo o frase.

(BBDD>registro>campos>párrafo/frase/subcampo)

Los campos pueden ser definidos como numéricos, de fecha y textuales. Al crear una base de datos que especifica que campos generarán índices y de que tipo (palabra, frase).

Operadores booleanos o lógicos: Se llaman booleanos porque se basan en el álgebra de Boole desarrollada por este matemático. George Boole desarrolló los métodos para combinar conjuntos utilizando expresiones algebraicas, que dieron como resultado la lógica booleana. El algebra de Boole se basa en tres operaciones lógicas: suma lógica (OR) ; producto lógico (AND) y diferencia lógica (NOT).

**O, OR** (Unión, suma lógica) → relaciona términos que corresponden a un mismo concepto. El OR exige que uno de los términos o ambos aparezca en un mismo registro, es decir, busca en todos los campos. En lenguaje común se expresaría como “uno de ellos o ambos”. (es útil para agrupar sinónimos. Ej: descuidado or abandonado or desaliñado or sucio)

**Y, AND** (Intersección, producto lógico) → Relaciona conceptos que deben estar presentes simultáneamente. AND exige que los términos que enlaza figuren simultáneamente en el registro. Al exigir que estén los dos términos se reduce el tamaño del resultado de búsqueda.

**NO, NOT** (Negación o diferencia lógica) → Excluye términos que no se desean. Exige que la palabra introducida tras el NOT no figure. Ejemplo: profesionales

---

<sup>1</sup> Codina, Lluís. Sistemes d'informació documental. Barcelona : Pòrtic, 1993

not técnicos. Excluye los documentos en que aparezca “técnicos”. Por ello debe manejarse con prudencia.

XOR (Pero no) → Uno de los términos está en el documento pero no ambos

XOR = OR – AND

OR = XOR + AND

AND= OR - XOR

NOT (no tiene formulación)

NOTA: En las bases de datos con interface de búsqueda por campos funciona del mismo modo pero limitado al campo concreto donde te sitúes.

Operadores posicionales o de proximidad: (aumentan la precisión de la búsqueda)

WITH, SAME, ADJ, NEAR

Estos operadores no sólo requieren que los términos estén presentes en el mismo documento, sino también que ocupen una relación de posición mutua determinada. (Busca términos en el mismo campo)

**SAME:** Exige que las palabras clave aparezcan en el mismo campo. Los documentos recuperados contienen todos ellos las palabras buscadas en el mismo campo, si bien el orden en que se encuentran no se tiene en cuenta. (se salta el delimitador de frase, ejemplo buscando en campo identificadores a dos autores de un mismo artículo)

**WITH:** Exige que las palabras buscadas aparezcan en la misma frase o sentencia del campo. Tampoco importa el orden en que se encuentran.

**NEAR:** Exige que las palabras buscadas aparezcan una junto a otra en la misma frase, pero sin tener en cuenta el orden. (recuperación near información). Este operador puede modificarse con un número calificativo que especifica lo cerca que han de estar las palabras. NEAR3 significa que las palabras deben estar separadas por un máximo de tres palabras. No tiene en cuenta el orden de las palabras.

**ADJ:** Exige que las palabras aparezcan una junto a la otra, en la misma frase y en el orden especificado. (gastos adj operativos). También puede calificarse con números pero a diferencia del near exige que las palabras aparezcan en el orden indicado.

Operadores de truncamiento: con los caracteres \* (ó \$ en algunos sistemas como BRS) y ¿ se realizan distintos tipos de truncamiento.

Truncamiento por la derecha: permite recuperar todas las palabras que empiezan con la misma cadena de caracteres, el símbolo \* se coloca inmediatamente después de la cadena del término introducido, sin ningún espacio en blanco de separación.

Ejemplo: metano\*

El sistema busca todos los elementos que empiezan por las letras metano, recuperando “metano”, “metanos”, “metanol”, “metanosulfonas”. Se emplea para simplificar la recuperación de variaciones sobre las palabras escogidas y para asegurar la recuperación de términos en singular y plural.

Truncamiento por la izquierda: Permite recuperar todas las palabras que terminan en una misma cadena de caracteres. El símbolo \* se escribe inmediatamente antes de la cadena a buscar, sin espacios en blanco.

Ejemplo: \*acetato

Esta búsqueda recupera registros en los que figuran palabras como “etilacetato”, “poliacetato”, etc. Se usa para buscar prefijos antepuestos a un término (o también a un número).

Truncamiento bilateral: algunos sistemas permiten truncar por los dos extremos de un núcleo común a recuperar con independencia de los prefijos y sufijos.

Ejemplo: \*esteroid\*

Esta búsqueda recupera términos como “esteroides”, “esteroideo”, “corticosteroides” y otras variaciones.

Truncamiento interno: se utiliza para localizar términos en los que el elemento variable está entre un prefijo y un sufijo comunes. Basta con sustituir los elementos variables con \* o ¿.

Ejemplo: inmuno\*químico

Recupera todos los términos que empiecen por “inmuno” y terminen por “químicos”, como “inmunoquímico”, “inmunohistoquímico”, etc.

El ¿ se emplea cuando existen dudas sobre la grafía exacta de un término en un solo carácter .

Ejemplo: Me?ico (para Mexico or Mejico)

Anidamiento de búsqueda, o búsqueda anidada: Las búsquedas pueden refinarse empleando paréntesis para “anidar” palabras clave e invalidar la secuencia normal de procesamiento de operadores. La secuencia normal es primero los de proximidad antes que los booleanos)

Una búsqueda sobre recuperación de información o datos podría construirse como sigue:

(información or datos) same recuperación

Los términos encerrados entre paréntesis son siempre los que primero se procesan. Los términos “información” o “datos” deben encontrarse en el mismo campo que “recuperación”.

Si se omiten los paréntesis, entraría en funcionamiento la secuencia normal y same se procesaría antes que or. El resultado será un número mayor de documentos. Se recuperarían documentos en los cuales aparece el término “datos” y “recuperación” en el mismo párrafo, más todos aquellos que contienen la palabra información.

Una alternativa al anidamiento es combinar los números secuenciales que dan los sistemas para cada búsqueda realizada.

Ejemplo:

- 1.- información or datos
- 2.- 1 same recuperación

Operadores relacionales o de comparación: > mayor que, < menor que, >= mayor o igual, <= menor o igual, igual, para campos numéricos de fechas.

Operadores de configuración (set) que permiten ordenar los registros de un modo determinado, seleccionar el operador por defecto, limitar el número de términos en la visualización de índices, visualizar el número de “ocurrencias”...

Otras posibilidades que ofrecen algunas bases de datos para recuperar la información:

delimitar por campos. Realiza la búsqueda solo en el campo o campos indicados.

Ejemplo: motores.tito,cont. ó In ti,abs: motores (depende del sistema)

Visualización de índices. Es una opción de gran valor para la recuperación de información y que algunas bases de datos permiten.

#### **4.- Etapas de la búsqueda documental**

4.1.- Conocer la necesidad de información: La entrevista de referencia (o la recepción de la consulta) *¿Qué se os viene a la cabeza cuando se habla de entrevista?*

(Merlo) La mejor manera de conocer la necesidad de información de los usuarios es la entrevista de referencia; se trata de la conversación que el personal referencista mantiene con el usuario con la intención de **delimitar y**

clarificar las demandas de información de éste. En esta entrevista el referencista va anotando las distintas facetas del tema de interés para que la respuesta se ajuste al tema de la consulta. El grado de profundidad de la entrevista de referencia dependerá de la consulta. Cuanto más especializado sea el tema planteado por el usuario más detallada será la entrevista.

Una entrevista de referencia completa deberá incluir las siguientes cuestiones:

-Tema principal de la búsqueda y temas relacionados con el mismo. El tema debe ser delimitado al máximo, excluyendo todo aquello sobre lo que el usuario no muestre interés. Es muy común que el usuario no tenga una idea clara acerca de qué es lo que le interesa y qué no, por lo que el personal debe hacer todas las preguntas necesarias hasta que la consulta esté perfectamente especificada.

-Información que ya conoce el usuario. Es importante saber qué datos, obras o autores ha manejado el usuario, ya que de lo contrario se puede dar información redundante.

-Objeto de la demanda o finalidad de la búsqueda: Saber para qué quiere el usuario la información será útil para determinar el grado de especialización y en qué fuentes se realizará la búsqueda.

-Profundidad o nivel de la información. Hay que preguntar que tipo de información va a ser útil para el usuario, a fin de eliminar aquella que por defecto o por exceso no se ajuste a las necesidades del usuario, ya sea por su nivel científico, por la exhaustividad de los datos, el grado de especialización o por cualquier otro motivo.

-Cantidad de información. En este punto se demandará sobre el número de respuestas que quiere el usuario. En ocasiones simplemente bastará con un número limitado de documentos, mientras que otras veces se necesitarán todas las referencias o datos que se encuentren.

-Idiomas de la información. En las búsquedas que se realicen aparecerán documentos en varias lenguas, por lo que habrá que preguntar al usuario qué idiomas quiere y cuáles no, a fin de eliminar aquellas referencias que no va a emplear.

-Límites cronológicos. Dependiendo del tema de la búsqueda la fecha de la información será útil o no, por lo que habrá que establecer a partir de que año quiere el usuario que se haga la búsqueda.

-Límites geográficos. (añadido mío) Determinar si el tema de su interés debe delimitarse solo a un ámbito geográfico concreto o no.

-Formato de las respuestas y modo de recepción. Preguntar si la respuesta se entregará en papel o en disquete y si el usuario lo recogerá personalmente o se le enviará por correo postal, fax o correo electrónico.

-Importe económico. Muchas consultas se resuelven únicamente en bases de datos o en servicios de información de pago, por ello hay que preguntar al usuario cuánto está dispuesto a gastar para la resolución de su consulta.

-Tiempo de respuesta. Urgencia del usuario por recibir la información y disponibilidad del centro para entregarla.

En la mayoría de estos apartados pueden ser empleadas dos tipos de preguntas: abiertas y cerradas. Las abiertas son aquellas en las que el usuario puede expresarse libremente para comunicar sus necesidades de información;

mientras que en las cerradas únicamente podrá contestar a una de las opciones que plantee como alternativa el personal referencista.

Ejemplos preguntas abiertas: ¿Qué tema le interesa?, ¿Qué autores conoce?, ¿Para qué necesita la información?. ¿Qué tipo de documento le interesa?

Ejemplos de preguntas cerradas: Objeto de la consulta (personal/estudios/profesional), Número de referencias (1 a 10/ 10 a 50/ más de 50), Forma de recepción (centro/envío postal/correo electrónico)

#### 4.2.- Metodología para establecer una estrategia de búsqueda (clarificación de la consulta)<sup>2</sup>

Una vez recogida la consulta a través de la entrevista de referencia se deberá proceder a su resolución. Se pueden marcar una serie de etapas:

-Análisis de la consulta:

Determinación exacta de la pregunta formulada por el usuario.

Análisis de los conceptos incluidos en la demanda.

Dividir la demanda en componentes lógicos.

Identificar los conceptos centrales.

Excluir conceptos innecesarios.

Cotejar resultados con el solicitante.

-Traducción de la consulta a términos de búsqueda:

Selección de los términos que reflejen los conceptos a buscar.

Traducir estos conceptos en tantos términos del lenguaje natural como sea necesario.

Elección de los descriptores, encabezamientos de materia, clasificaciones, etc. que se correspondan con los conceptos a buscar.

Decidir el grado de exhaustividad y especificidad de la búsqueda. La exhaustividad de la búsqueda es la medida en la que todas las facetas, aspectos o conceptos identificados en la solicitud de información deben estar presentes en la estrategia de búsqueda. Especificidad es el nivel de detalle con el que una determinada faceta debe estar representada. La especificidad depende directamente del vocabulario que se utilice. Confeccionar el perfil de búsqueda o conjunto de términos del lenguaje natural, palabras clave o descriptores que van a ser utilizados para la realización de la búsqueda.

Herramientas para la selección de términos: información proporcionada por el solicitante, diccionarios, tesauros, clasificaciones, encabezamientos de materias, listas de frecuencia de términos, obras de referencia.

Comprobar que se utilizan todas las formas necesarias de una palabra, que se incluyan los sinónimos necesarios y los descriptores correctos y que se hayan explotado adecuadamente las relaciones entre los términos.

-Selección de las fuentes de información: Elección de las bibliografías, obras de referencia, bases de datos, etc. que se considere más apropiados para resolver la demanda planteada.

---

<sup>2</sup> Amat, Nuria. *La biblioteca*. Barcelona : Scripta, 1982



-Traducción de los términos de búsqueda al lenguaje documental de cada fuente:

Comparar los descriptores, materias, palabras clave, encabezamientos, etc. obtenidos previamente (fase traducción de la consulta a términos de búsqueda) con los empleados en la fuente de información que se está consultando y emplear los términos que más se ajusten.

-Establecer la ecuación de búsqueda y posibles ecuaciones alternativas (también llamada esta etapa construcción de la *estrategia* o perfil de búsqueda): Relacionar los términos mediante los operadores booleanos. Contemplar otros posibles operadores (proximidad, truncamiento, etc.) en función de los términos de búsqueda y los aspectos de interés.

Emplear los paréntesis para anidar búsquedas.

Conviene hacer la ecuación de búsqueda en varios pasos, siguiendo una progresión lógica de creación de conjuntos.

-Llevar a cabo la búsqueda y visualizar los resultados: Comprobar si los documentos resultantes son relevantes. En caso contrario hay que modificar la estrategia y repetir la búsqueda.

-Síntesis de las respuestas encontradas: Seleccionar y organizar la información válida para la resolución de la demanda, eliminando la superflua y la que no se ajusta a la solicitud.

-Preparación del documento respuesta:

Elegir el formato de salida antes de imprimir o descargar los resultados para la entrega al usuario.

(Relevancia: el resultado se ajusta a la búsqueda. Grado en que el documento cubre la solicitud de información.

Pertinencia: Adecuación del resultado respecto a los usuarios a los que va dirigida la información. Puede ser relevante y no interesar al usuario)

## BIBLIOGRAFIA

AENOR. UNE 50/113 Documentación e información. Vocabulario. Parte 1. Conceptos fundamentales.

Baquedano Alcocer, Arrate ; Pastor López, Oscar. *Bases de datos documentales: cómo estructurar la información no estructurada*. Valencia : Diputación de Valencia, 1993

Caridad Sebastián, Mercedes. Recuperación. En: *Operaciones de la cadena documental*. Madrid : Instituto Oficial de Radio y Televisión, 1988

Codina, Lluís *Sistemes d'informació documental*. Barcelona : Pòrtic, 1993

Fuentes i Pujol, M Eulalia. *Documentació, telecomunicacions i informàtica: la teledocumentació*. Barcelona : Pòrtic, 1990

Guinchat, Claire ; Menou, Michel. La búsqueda de información. En: *Introducción general a las ciencias y técnicas de la información y documentación*. Madrid : CINDOC, 1992

Merlo Vega, José Antonio. El servicio bibliotecario de referencia. *Anales de documentación*. N° 3, 2000, p. 93-126

Pérez Alvarez-Ossorio, J. R. Almacenamiento y recuperación de información. En: *Introducción a la información y documentación científica*. Madrid : Alhambra, 1988

Salvador Oliván, José Antonio ; Angós Ullate, José María. *Técnicas de recuperación de información. Aplicación con Dialog*. Gijón : Trea, 2000