

BÚSQUEDA Y RECUPERACIÓN DE INFORMACIÓN

TEMA 3

Tema 3

Sistemas de Rec. de Inf.

- La **Recuperación de Información** es un área de conocimiento que se ocupa de la recopilación, representación, almacenamiento, organización y acceso a recursos de información como documentos, páginas web, catálogos online, registros estructurados o semiestructurados u objetos multimedia. La representación y organización de los ítems de información debe ser tal que proporcione a los usuarios acceso fácil a la información de su interés (Baeza Yates, R.; Ribeiro-Neto, B. Modern information retrieval, p. 1).

Tema 3

Sistemas de Rec. de Inf.

- Los **Sistemas de Recuperación de Información (SRI)** son programas informáticos que sirven para gestionar el proceso de indexación y búsqueda de información en catálogos, bases de datos y buscadores. Cuentan con diversos componentes o subsistemas, por ejemplo el de introducción de datos o el subsistema de consultas, para la gestión y el procesamiento de grandes cantidades de información. Los SRI utilizan una lógica determinada para conectar la pregunta de los usuarios con la respuesta más adecuada posible a su demanda (Cid Leal, P. Cómo y dónde buscar fuentes de información, p. 23).
- El conocimiento de la estructura interna de un SRI nos ayudará a conocer mejor el procedimiento que se sigue para recuperar la información en una base de datos o en un buscador de Internet.

Tema 3

Sistemas de Rec. de Inf.

- La arquitectura de un SRI consta de varios componentes, destacándose los siguientes:
 - Componente de Recopilación de la colección, que también puede estar automatizada en el caso de un motor de búsqueda como Google, p. ej.
 - Componente de Indexación. La estructura del índice más empleada es la denominada índice inverso, compuesto por todas las palabras distintas que figuran en los documentos de la colección y, para cada una de ellas, una lista de los documentos que la contienen. Este componente se ocupa, en resumen, de la representación interna de los documentos para llevar a cabo una recuperación rápida y eficaz.
 - Componente de representación de las necesidades de información de los usuarios. Este componente convierte dichas necesidades en consultas que el sistema puede entender y procesar.
 - Componente de recuperación, consistente en equiparar las consultas introducidas al sistema con la representación interna de los documentos de la colección para extraer los ítems de información considerados relevantes para solventar la necesidad informativa del usuario, y en su caso ordenarlos por nivel de pertinencia en relación a la necesidad informativa.

Tema 3 Estrategias de búsqueda y rec.

- Debe distinguirse entre la búsqueda de información y la recuperación de información. Como hemos visto ya, la recuperación de información es un proceso técnico que involucra un tratamiento automatizado de los documentos de la colección.
- La **búsqueda de información** es un proceso mucho más amplio, que va desde que surge la consciencia de carencias que impiden realizar una tarea hasta la adquisición de la información que contribuye a superar dichas carencias y afrontar o culminar la tarea propuesta inicialmente.

Tema 3 Estrategias de búsqueda y rec.

- En el proceso de búsqueda de información destacan los siguientes pasos importantes:
 - Delimitación y concreción de la necesidad informativa. En muchas ocasiones no sabemos especificar qué información concreta podría sernos de utilidad, o ignoramos cuál es la denominación adecuada de lo que buscamos.
 - Delimitación de la o las fuentes de información más adecuadas a la necesidad de información planteada.
 - Formulación de la/s consulta/s en la fuente de información. Este paso incluye, en su caso, el conocimiento del funcionamiento de la fuente de información, sobre todo si esa fuente de información incluye un sistema de recuperación de información (catálogo, base de datos, Internet).
 - Evaluación del resultado obtenido, que puede ser positiva o negativa. Si es positiva, el proceso de búsqueda se da por concluido. Si la evaluación es negativa, el proceso vuelve a retomarse en alguno de los puntos anteriores.

Tema 3 Estrategias de búsqueda y rec.

- Se entiende por **estrategia de búsqueda** el conjunto de decisiones adoptadas por un usuario en relación a cada uno de los pasos del proceso de búsqueda (concreción de la necesidad, elección de la fuente de información, formulación de la consulta y evaluación del resultado).
- Dado el carácter iterativo de la búsqueda de información, es habitual que el proceso incluya sucesivas **modificaciones en la estrategia de búsqueda**.
- En consecuencia, La búsqueda de información es un **proceso iterativo** en el que se va modificando la estrategia de búsqueda hasta delimitar correctamente la necesidad, las fuentes más adecuadas y las consultas pertinentes que nos conduzcan a la información útil deseada.

Tema 3 Estrategias de búsqueda y rec.

- En relación a la etapa de formulación de la consulta en un proceso de búsqueda de información, los SRI que incorporan catálogos, bases de datos y buscadores de Internet suelen incluir dos opciones de búsqueda:
 - Búsqueda básica: Introducción de la consulta directamente en un área de texto en blanco que proporciona el SRI.
 - Búsqueda avanzada: Formulario que permite al usuario especificar diversos aspectos de la consulta (términos, delimitación geográfica, cronológica...)

Tema 3 Estrategias de búsqueda y rec.

- De igual forma, aunque cada SRI puede tener un procedimiento propio de interrogación, los catálogos, bases de datos y buscadores habitualmente comparten un conjunto de operadores comunes, pudiendo variar en mayor medida su expresión formal.
- Entre los operadores comunes (aunque no todos los SRI disponen de todos ellos simultáneamente), destacan los siguientes:
 - Operadores booleanos
 - Operadores de proximidad
 - Operadores de truncamiento
 - Operadores de campo

Tema 3

Operadores booleanos

- **UNIÓN:** Con este operador entre dos términos, se obtienen los documentos de la colección que contienen al menos uno de dichos términos. Ej: pera OR manzana
- **INTERSECCIÓN:** Con este operador entre dos términos, se obtienen los documentos de la colección que contienen simultáneamente ambos términos. Ej: caballos AND carreras
- **EXCLUSIÓN:** Con este operador adscrito a un término, se obtienen los documentos de la colección que NO contienen dicho término.
Ej: NOT dulce

Tema 3

Operadores de proximidad

- OPERADOR DE FRASE: Indica que el conjunto de palabras o términos deben aparecer exactamente en el orden señalado. Ej: “violencia de género”
- OPERADOR DE DISTANCIA: Indica que los términos aparezcan separados un máximo de n palabras entre sí.

Ej: fuentes (1) información

[fuentes de información, fuentes con información,... LOS RECUPERA]

[Las fuentes de La Granja ofrecen información valiosa... NO LO REC.]

T.3 Operadores de truncamiento y de campo

- **OPERADORES DE TRUNCAMIENTO:** Permiten sustituir caracteres de las palabras introducidas en las consultas, ampliando de esta forma el número de palabras que el SRI recupera.

Ej: bibliotec* [recupera los términos biblioteca, bibliotecas, bibliotecario, bibliotecaria, bibliotecarios, bibliotecarias, biblioteconomía, bibliotecología...]

- **OPERADORES DE CAMPO:** Permiten especificar el campo (título, autor, tema, año de publicación...) en el que deben figurar las palabras introducidas en las consultas.

Ej: ti:bibliotec*; au:"lope de vega"; su:renacimiento