

3. Los principios básicos de recuperación de información

Se explica el modelo básico de proceso de búsqueda, así como la formación de ecuaciones y el uso de operadores.

- [3.1. El proceso de búsqueda](#)
- [3.2. Ecuaciones y operadores](#)
- [3.3. Interpretación de resultados. Límites](#)

3.1. El proceso de búsqueda

Los procesos de búsqueda de información distan de ser perfectos y pautables. Cada proceso es diferente, y se ve afectado por numerosos factores, tanto internos del propio proceso y de la herramienta utilizada, como externos, relacionados con el propio usuario y con sus contextos social, educativo o técnico. La necesidad de respuestas rápidas, o el nivel de profundidad deseado, también influyen en las decisiones que el usuario toma en cada momento. No debe olvidarse que, aunque se trata de un proceso mediado tecnológicamente, **es el usuario el que toma decisiones que afectan al propio proceso.** Finalmente, la satisfacción que el usuario siente o expresa con el proceso también es dependiente de factores intrínsecos al propio usuario.

Un aspecto que es necesario destacar antes de iniciar un proceso de búsqueda es el **problema que puede plantear la transferencia entre tres niveles de expresión:**

1. En **primer lugar**, el usuario tiene un **modelo mental de su carencia o necesidad de información**, pero es necesario que **traslade ese modelo al lenguaje formal**. Esta primera transformación ya puede suponer un riesgo de pérdida de elementos.
2. En **segundo**, el usuario **transformará la expresión en lenguaje formal en una expresión en el lenguaje de búsqueda** que utiliza el motor (la clásica expresión o ecuación de búsqueda).
3. **Finalmente**, el **motor transformará esa expresión de acuerdo al algoritmo que utilice internamente** para representar la información. Las respuestas que ofrezca responderán a esta última transformación, y podrán, a su vez, responder o no a la necesidad original del usuario.

Los siguientes párrafos recogen una propuesta clásica de un proceso de recuperación; pero debe recordarse siempre que la búsqueda y recuperación de información no son procesos y actividades exactas: pueden encontrarse varios procesos y soluciones distintas para el mismo problema:

1. **Planteamiento del tema y nivel de conocimientos:** Este trabajo permite definir con claridad cuál es la necesidad de información del usuario. Debe establecer claramente cual es el objetivo de interés, así como los límites del mismo. Hay objetivos que pueden parecer adecuados, pero que en realidad necesiten un refinamiento; deben plantearse todas las posibles situaciones que puedan darse, y pensar en varias aproximaciones al problema. Se debe establecer cual es el propio nivel de conocimientos sobre el tema. Si el nivel de conocimientos es adecuado, se podrá abordar el problema de la fiabilidad con mayores garantías. En cualquier situación, esta fase debe dar como resultado una formulación clara e inequívoca del objetivo de su búsqueda.
2. **Identificación de los tipos de información:** El web contiene diferentes tipos de información, tanto por el tipo de fichero que los contiene, como por el objetivo y finalidad de las páginas web y de los creadores de las mismas. Es necesario establecer la posible utilidad de cada uno de estos tipos de documentos, y no desdeñar ninguno a priori, ya

que por exploración se puede encontrar información complementaria que resulte pertinente.

3. **Selección de los recursos de información y de las herramientas de consulta:** La selección de los recursos de información, es decir, motores de búsqueda genéricos, recursos de información especializada, etc. a utilizar, es de suma importancia. Dependiendo del contexto de actividad del usuario, en muchas ocasiones el conocimiento de estas fuentes especializadas es directo, sin necesidad de acudir a una búsqueda previa en un motor generalista.

4. **Ejecución de las expresiones o ecuaciones de búsqueda, evaluación e iteración:** Se trata de la formulación y ejecución de las expresiones o ecuaciones de búsqueda en el motor o recursos que se trate (correspondería con las segunda y tercera transformaciones indicadas previamente). Es un proceso iterativo: recepción de respuestas, valoración

nueva ejecución, nuevas respuestas, [detalle en el apartado siguiente](#)).

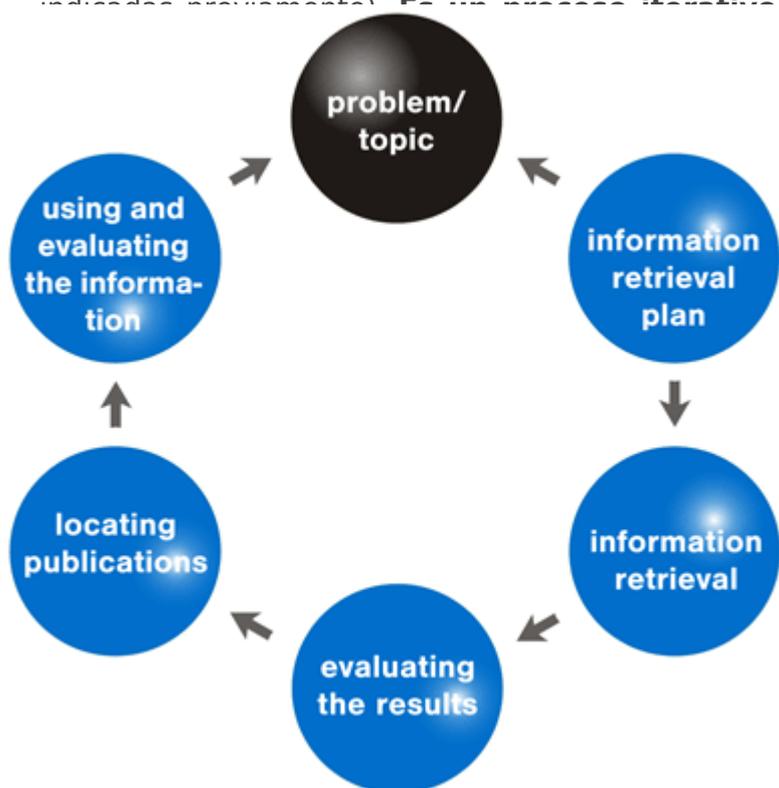


Fig. 8. Proceso genérico de búsqueda de

información ([fuente original](#)).

El proceso acabará cuando el usuario considere que su necesidad de información puede ser resuelta con los documentos y la información obtenidos. Después del uso de motor de búsqueda, deberán realizarse la tareas de acceso a los originales, valoración de contenido, selección y uso en virtud de la necesidad original.

Financiado por el Ministerio de Educación y Formación Profesional y por la Unión Europea - NextGenerationEU



3.2. Ecuaciones y operadores

La operativa más conocida en la búsqueda de información es la formulación de expresiones o ecuaciones de búsqueda. Como se ha explicado en el apartado anterior, en un proceso bien planteado ésta debería ser una fase más, y debería ir precedida de una adecuada planificación. En la actualidad, los usuarios suelen adoptar un enfoque de “*escribir y esperar*”: escriben en la interfaz de consulta una palabra o frase que consideran adecuada, y esperan los resultados. Suelen evaluar los 10 primeros, o lo que muestra la primera página de resultados, y en pocas ocasiones avanzan más allá. Tampoco suelen acudir al uso de operadores para mejorar los resultados, ya que normalmente cambian los términos o frase usada, con la esperanza de obtener mejores resultados. El proceso de interrogación suele organizarse alrededor de tres tareas principales:

1. **Formulación de la ecuación, ejecución y recepción de respuestas:** En ella se introduce en la interfaz la expresión que reúne los términos elegidos, y los operadores que establecen las relaciones existentes entre aquellos. Los motores ofrecen siempre páginas de ayuda, en las que explican las posibilidades del lenguaje de interrogación que usan, y suelen incluir ejemplos. También ofrecen interfaces simples y avanzadas para formular las ecuaciones. La recomendación sería utilizar las interfaces avanzadas, ya que van a ofrecer más potencial y parámetros que ayudan a perfeccionar las ecuaciones y a obtener resultados ajustados.
2. **Visualización y preselección de respuestas pertinentes:** Se trata de la exploración de las respuestas obtenidas y, en su caso, la consulta de los documentos originales. Los datos que ofrece el listado se utilizan para llevar a cabo una primera selección de aquellas respuestas que parecen relevantes, y para desechar aquellas que no sean adecuadas. Acudir a consultar el documento original exige establecer una conexión con el mismo, lo que lleva a un cambio de la interfaz de visualización y a la pérdida del listado de respuestas por efecto de la navegación. Para evitar esa pérdida de contexto, y evitar la sobrecarga cognitiva, es recomendable establecer la conexiones en ventanas o pestañas nuevas. Si el documento al que se ha accedido interesa, la recomendación es guardar el url de acceso en funcionalidades de marcadores o herramientas similares.
3. **Replanteamiento de estrategias:** Si tras analizar las quince o veinte primeras respuestas no se han obtenido resultados satisfactorios, es necesario modificar la búsqueda. El cambio puede referirse a las ecuaciones utilizadas, o al motor o herramienta seleccionado, o a ambas cosas. Si el número de respuestas obtenido es muy elevado, y los primeros resultados son poco pertinentes, muy generales, debe formularse una nueva ecuación de búsqueda, con más condiciones y limitaciones. En el caso contrario, con nulo o escaso número de resultados, puede suceder que a) si la ecuación no es restrictiva, entonces no hay documentos, o los documentos no contienen esos términos; b) que la

ecuación sea demasiado restrictiva, con demasiadas condiciones. Para estos casos hay que experimentar con una ecuación con menos condiciones, y usar términos sinónimos o similares a los usados anteriormente.

Los lenguajes de interrogación utilizan una estructura en la que participan órdenes, operadores y operandos, los cuales, combinados según un conjunto de reglas sintácticas establecidas en el propio lenguaje, permiten formular expresiones de interrogación que reciben el nombre común de ecuaciones. Los lenguajes de recuperación que utilizan los motores u otros recursos y fuentes de información están disponibles para el usuario a través de una interfaz propia, que limita las posibilidades de construcción de expresiones, aunque teóricamente sin perjuicio de la necesaria complejidad. En realidad, esta interfaz supone una simplificación de esas posibilidades, porque no suele mostrar todos los operadores disponibles. Los operadores disponibles para la recuperación de información en internet pueden clasificarse de la siguiente forma:

- **Operadores booleanos:** todos los motores admiten los tres operadores booleanos básicos (Y, O, NO - AND, OR, NOT). Esto permite construir ecuaciones basadas en la lógica booleana, y la resolución de las mismas mediante la aplicación de la teoría de conjuntos. Al igual que en los sistemas clásicos de recuperación de información, los operadores booleanos trabajan considerando exclusivamente la presencia o ausencia en los documentos de los términos incluidos en la ecuación. Algunos motores admiten los llamados "pseudobooleanos", con los símbolos +/-, que obligan a la presencia o ausencia del término al que preceden.

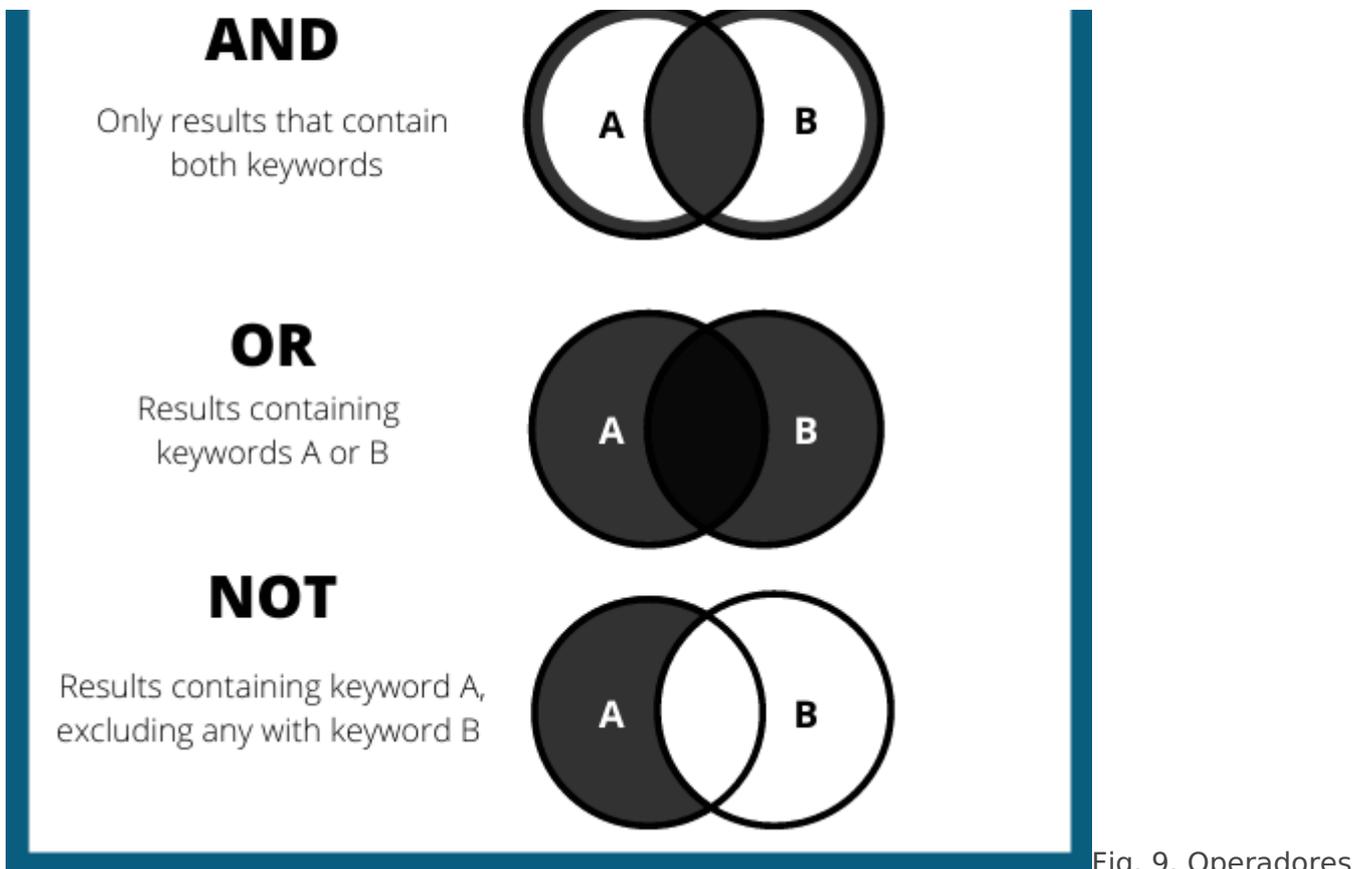


Fig. 9. Operadores

booleanos de conjuntos ([fuente original](#)).

- **Operadores de frase o expresión:** de la misma forma, casi todos admiten el uso de frase o expresión exacta, generalmente entrecomillada, o mediante la activación de una casilla o botón específico para tal fin. En realidad, se trata de la utilización de un operador de posición relativa, el de adyacencia, con la limitación de respetar el orden de aparición de los términos.
- **Operadores de límite:** dentro de lo que los motores identifican en sus interfaces como límites, tienen cabida un dispar conjunto de operadores, que varían notablemente de unos motores a otros. Dentro de estos límites es posible identificar, por ejemplo, operadores de rango (relacionados con fechas, por ejemplo), o truncamientos (palabras de misma raíz, búsqueda por lema o raíz de los términos).
- **Operadores de posición o campo:** algunos motores incorporan la posibilidad de utilizar el contenido de determinadas etiquetas HTML, presentes en los documentos, como campos específicos, de modo que pasan a actuar como operadores de posición absoluta. Si los términos buscados aparecen dentro de un título, un enlace u otra etiqueta significativa, se deduce que el contenido del documento es más relevante.
- **Operadores de relación entre documentos:** una posibilidad de sumo interés es la disponibilidad de operadores que emplean los enlaces hipertextuales entre documentos presentes en los mismos, para recuperar documentos enlazados, tanto en destino, como

en origen. La utilización de los mismos nos permite realizar búsquedas por exploración, tras localizar un documento que actúa como punto de partida o “semilla”.

- **Operadores de ejemplo:** bajo este nombre se agruparían los algoritmos que hacen posible localizar páginas o recursos web que ofrecen un alto nivel de similaridad con un ejemplo o patrón que ha sido identificado por el usuario como altamente relevante. El ejemplo clásico sería la funcionalidad “Más como éste” disponibles en los motores de búsqueda.

<https://www.youtube.com/embed/PrJl7kYdSc8>

Videotutorial: Estrategias de búsqueda de información (2020)

Financiado por el Ministerio de Educación y Formación Profesional y por la Unión Europea - NextGenerationEU



3.3. Interpretación de resultados. Límites

En primer lugar, debe tenerse en consideración que puede existir cierto nivel de disfunción entre el contenido informativo real de una página web, en virtud de las características de los documentos digitales, que pueden cambiar de manera dinámica muy rápidamente, y la representación que se obtiene de ella mediante los procesos de indización automática que llevan a cabo los motores.

Una vez considerado esto, y **tras llevar a cabo una búsqueda, es necesario analizar la pertinencia y relevancia de los resultados obtenidos**. Una alta pertinencia se caracteriza por la recuperación de los documentos adecuados a la información que se deseaba obtener, evitando la introducción de otros que puedan distorsionar el conjunto, impidiendo una correcta recuperación de la información. En sistemas de tratamiento y recuperación de información documental, ésta debe ser exacta, exhaustiva, precisa, oportuna, íntegra y significativa. Por contra, el **silencio** (falta de documentos) y el **ruido** (exceso de documentos, muchos de ellos no significativos) caracterizan a un bajo nivel de pertinencia. Debe tenerse en cuenta que la pertinencia hace referencia a la correspondencia entre la ecuación formulada y los resultados obtenidos, lo que quiere decir que un resultado puede ser pertinente... pero no ser relevante para el usuario, ya que éste puede conocer su contenido previamente, o no ajustarse con exactitud a lo que necesita. **Será el usuario final, durante la evaluación de las respuestas, el que decidirá sobre la relevancia o no de un resultado o de una información para resolver la necesidad que tenía planteada.**

<https://www.youtube.com/embed/olvc4OPHzno>

Videotutorial: Criterios de búsqueda de información en internet (2020)

Como resultado del análisis anterior, cabe plantearse la adecuación de la respuesta obtenida a lo esperado. Pueden darse dos situaciones, ante las cuales deben adoptarse diferentes acciones. En un primer caso, puede suceder que el resultado de la ecuación sea demasiado escaso. Entonces procede ampliarla, lo cual puede hacerse con la utilización de términos más genéricos, sinónimos, o ampliación de truncamientos. En un segundo caso, si el resultado obtenido es excesivamente amplio, deben utilizarse medidas contrarias: utilización de términos más específicos, reducción de truncamientos, etc.

Por último, en entornos digitales puede producirse un problema que afecta a los procesos de búsqueda. Al trabajar en un entorno bidimensional a través de la interfaz de un navegador, la consulta puede quedar atrás al acudir a consultar una respuesta. La propia consulta de la

respuesta puede llevar a iniciar un proceso de exploración de otras páginas o documentos web. En un momento **puede producirse una desorientación cognitiva**, en la cual el usuario pierde el foco de referencia (la búsqueda original) y se ve inmerso en una navegación de continuos saltos adelante y atrás. Es necesario prever ese potencial problema, y adoptar pautas de navegación que limiten las complicaciones que puedan surgir durante el proceso de comprobación y selección de resultados.

Financiado por el Ministerio de Educación y Formación Profesional y por la Unión Europea - NextGenerationEU

