



## Módulo 1

### 1.3 EL PROCESO DE BÚSQUEDA

Por **María Dolores Olvera Lobo**

Catedrática del Departamento de Información y Comunicación. Grupo HUM-466 Información Científica. Universidad de Granada.

---

#### 1. EL PROCESO DE RECUPERACIÓN DE INFORMACIÓN

El concepto "sociedad de la información", tantas veces utilizado, se refiere a aquella en la que las tecnologías cobran un importante protagonismo, no sólo facilitando la producción y difusión de información sino también todo tipo de actividades (sociales, culturales y económicas, entre otras). La explosión de información en la que estamos inmersos, tanto en el ámbito profesional como en el cotidiano, hace imprescindible que incorporemos nuevas formas de encarar esta realidad y, sobre todo, que conozcamos cuáles son y cómo se utilizan las herramientas existentes para acceder y gestionar la gran cantidad de datos que nos rodea.

La recuperación de información ayuda a alcanzar este cometido. Consiste en el conjunto de tareas mediante las cuales los usuarios, es decir, aquellas personas con necesidades informativas localizan y acceden a los recursos de información que son pertinentes para la resolución del problema planteado. Para desarrollar este proceso los usuarios debemos cumplir varias etapas:

1. Definir nuestra necesidad de información.
2. Seleccionar la herramienta o sistema de recuperación de información más adecuado para satisfacer dicha necesidad de información.
3. Trasladar nuestra necesidad de información al lenguaje documental (este aspecto se detalla más adelante) propio del sistema que se utilizará en ese caso. No obstante, hay sistemas de recuperación de información en los que no se utiliza ningún tipo de vocabulario controlado (como los buscadores web o numerosas bases de datos documentales de texto completo), en cuyo caso resultará necesario precisar enormemente la consulta (utilizar sinónimos o expresar los conceptos en varias lenguas, entre otras estrategias).
4. Aplicar el lenguaje de interrogación o lenguaje de consulta propio de cada herramienta de recuperación. En esta etapa, se ejecuta la consulta en el sistema mediante la ecuación de búsqueda generada.
5. Consultar los resultados obtenidos con el fin de determinar si resultan relevantes (es decir, adecuados) o no a nuestra necesidad de información.



6. Replantear y modificar la consulta si los resultados obtenidos no nos resultan relevantes.
7. Seleccionar y obtener los documentos (es decir, cualquier recurso informativo incluido en la base de datos del sistema tales como páginas web, artículos de publicaciones periódicas, mapas, material audiovisual, entre otros).
8. Almacenar los resultados, si procede.

La respuesta “ideal” de un sistema de recuperación de información (SRI), está formada solamente por documentos relevantes a la consulta, generalmente ordenados en un ranking de resultados. No obstante, en la práctica ocurre que también pueden aparecer resultados no deseados. Esto se debe a que, entre otros motivos, existe el problema de compatibilizar la forma de expresar la necesidad de información, es decir, de plantear la consulta, y el lenguaje en el que está expresada dicha información en los documentos de la base de datos.

## 2. NECESIDADES DE INFORMACIÓN DE LOS USUARIOS

El proceso de recuperación de información en un entorno automatizado tiene, por tanto, una primera fase, en la que se debe definir claramente el problema informativo que se desea solucionar. Hay que tener en cuenta que, actualmente, por Internet, fluyen, ininterrumpidamente, datos, ficheros, textos, gráficos, imágenes, en definitiva, informaciones de todo tipo. Esta información cuenta con unas características peculiares entre las que se encuentran las siguientes:

- Inmensa cantidad de información: Aunque resulta prácticamente imposible cuantificar la información disponible de forma exacta, se sabe que la red permite acceder a más de un billón de servidores web.
- Diversidad de contenidos: Las oportunidades y los contenidos que ofrece la red para los investigadores, los profesionales, y los usuarios en general son múltiples y diversos. Hoy en día es posible realizar una tesis doctoral con, exclusivamente, la información extraída de Internet, o utilizar la red para llevar a cabo el trabajo de campo mediante encuestas, etc. Por otro lado, prácticamente todas las grandes bibliotecas o librerías mundiales están hoy accesibles a través de la red. De esta manera, es posible hallar casi cualquier libro que nos interese, comprobar su localización exacta, su estado, y solicitar su envío, préstamo, copia, etc. También podemos acceder directamente a grandes bases de datos, o fuentes de información generales, entre otros muchos ejemplos.
- Acceso ágil a la información: Destaca la notable rapidez y comodidad con la que es posible localizar aquellas fuentes o datos buscados por el usuario.
- Múltiples tipos de formatos: La aparición de nuevos tipos de información viene, en muchas ocasiones, de la mano de formatos específicos para su visualización.
- Información hipermedia, es decir, multimedia –que combina texto, sonido, imagen fija y en movimiento–, e hipertextual –que incluye links o enlaces a otros recursos–.



- La ubicuidad de la información y del usuario. La información no está estructurada, controlada, ni centralizada. Se encuentra dispersa en ordenadores de todo el mundo, con diferentes formatos y contenidos, sin normalización ni coordinación, por lo que hay duplicidades y omisiones.
- Dificultad de filtrar información. Esa extraordinaria abundancia de información puede tornarse un problema, que hace que se convierta en extremadamente ardua la tarea de filtrar y seleccionar la información realmente útil, más aún cuando se desconocen los límites y la existencia de contenidos potencialmente necesarios.
- Actualización y contrastación de la información: Otro aspecto a tener en cuenta es la verificación de la información y la constante necesidad de actualización. Ante una búsqueda de información, uno de los aspectos más importantes es la necesidad de cuestionarse hasta qué punto es válida y fiable. A ello se suma otra exigencia, la de encontrar aquella información que realmente se adapte a nuestras necesidades específicas.

Por tanto, para satisfacer nuestras necesidades de información es necesario usar las fuentes de información más convenientes, siempre teniendo en cuenta su calidad. Además, con el fin de obtener los resultados más adecuados, hemos de ser cuidadosos al expresar nuestra consulta al sistema de recuperación de información. Esta consulta se formula como una ecuación de búsqueda, adaptando dicha expresión al lenguaje del sistema.

### 3. EL PROBLEMA DE LOS LENGUAJES

En la recuperación de información entran en juego los llamados tres lenguajes: el lenguaje natural, el lenguaje documental y el lenguaje de interrogación. Como hemos indicado, en el proceso de recuperación de información, la primera etapa consiste en el planteamiento de nuestra necesidad de información (es decir, determinamos la información que deseamos obtener) mediante el lenguaje natural. El lenguaje natural es el lenguaje humano que utilizamos habitualmente para la comunicación y que, es enormemente rico en posibilidades. No obstante, precisamente por eso, también presenta problemas de ambigüedad para los sistemas de recuperación de información –los cuáles procesan de forma automática los documentos incluidos en su base de datos y las consultas de los usuarios–. De esta manera, la ambigüedad del lenguaje natural –como la polisemia o sinonimia– puede causar ciertas dificultades en la búsqueda y recuperación de información.

La necesidad de información expresada en lenguaje natural, no siempre es posible plantearla "tal cual" en un sistema de recuperación de información para realizar las consultas, es el caso, generalmente de las bases de datos bibliográficas. Algunos sistemas representan el contenido temático de los recursos a los que ofrecen acceso utilizando los denominados lenguajes documentales. Estos son las listas de encabezamientos de materias, las clasificaciones temáticas y los tesauros. Uno de los aspectos más destacados en recuperación de información es el problema planteado por la traslación o "traducción" de los conceptos de búsqueda a estos diferentes tipos de lenguajes, cada uno de ellos con características



propias. La principal característica de los lenguajes documentales es el control terminológico, que permite la representación de los conceptos de los que tratan los documentos sin ambigüedades, garantizando el empleo de los mismos términos para expresar los mismos conceptos. Por tanto, en el proceso de recuperación de información deberemos trasladar los conceptos que representan esa necesidad de información, y las relaciones entre ellos, al lenguaje documental que utilice el sistema que vamos a consultar.

Por ejemplo: si en una biblioteca queremos localizar manuales de gramática inglesa hemos de comprobar la forma en que están expresados esos conceptos en ese catálogo automatizado. Así, podemos encontrar que tal vez necesitemos tres términos, denominados encabezamientos de materias, para expresar esa consulta: manuales, gramática, lengua inglesa.

Una vez obtenidos y validados los términos que deban utilizarse, así como las relaciones entre ellos, las expresiones resultantes deben transformarse a un "tercer lenguaje", el cual cumple la función de interrogar al sistema de recuperación, de una forma consistente y comprensible para este, con la finalidad de localizar aquellos documentos que cumplan los requisitos establecidos. A este tercer lenguaje se le denomina lenguaje de consulta o lenguaje de interrogación. Partiendo de los términos y conceptos ya expresados por el usuario y, en su caso, validados por el lenguaje documental correspondiente, será preciso utilizar las prestaciones del sistema (operadores lógicos, operadores de proximidad, truncamiento, búsqueda por campos, entre otros) para expresar de manera formalizada la necesidad de información. El resultado de esta etapa es la denominada ecuación de búsqueda, una expresión de la necesidad de información capaz de ser entendida y procesada por el sistema que estamos consultando.

*Por ejemplo*, supongamos que nuestra necesidad de información es la siguiente: Queremos buscar recetas de pastel de manzana o de pera que no contenga huevo. Esta necesidad de información está expresada en lenguaje natural. Si bien los buscadores web no aplican ningún lenguaje normalizado o controlado (y, por tanto, podemos usar los términos de búsqueda como deseamos), al realizar la búsqueda en un catálogo de bibliotecas o en una base de datos bibliográfica especializada puede requerirse que expresemos esa necesidad de información mediante un lenguaje documental. Supongamos asimismo que estamos utilizando un catálogo web de bibliotecas, donde, para nuestro ejemplo, los encabezamientos de materia son los siguientes: recetarios, tartas, manzanas, peras, huevos

Ya hemos utilizado por tanto el segundo lenguaje implicado en la recuperación de información. Sin embargo, para plantear la consulta de forma comprensible para el sistema, es necesario además formularla mediante una ecuación de búsqueda que este pueda procesar. Para ello se seleccionarán las opciones de búsqueda adecuadas, lo que, en nuestro caso, podría dar lugar a la siguiente ecuación de búsqueda (si bien la forma exacta de expresarla cada sistema varía):

**MATERIAS:** ((*recetarios NEAR tartas*) AND (*manzanas OR peras*)) NOT huevos



#### 4. LOS TÉRMINOS DE BÚSQUEDA

Como venimos manteniendo, nuestra necesidad de información ha de ser expresada en forma de consulta al sistema. Los términos incluidos en la consulta se denominan palabras clave o términos de búsqueda, y juegan un papel fundamental en el proceso de recuperación puesto que representan la semántica, es decir, el significado de nuestra necesidad de información. Hay que tener en cuenta que cada palabra importa. En nuestra consulta debemos incluir términos significativos y específicos evitando las denominadas palabras vacías (es decir, aquellas con función gramatical o carentes de significado relevante para nuestra necesidad informativa). Por tanto, la consulta puede estar formada por uno o varios términos, o por frases literales. En ocasiones, el sistema tiene en cuenta el orden de las palabras en la consulta planteada mientras que, a veces, el uso de las mayúsculas en los términos de búsqueda y la puntuación que incluyamos suele ser ignorada.

Por ejemplo, ante la siguiente búsqueda, el proceso de selección de términos de búsqueda, de representación de conceptos, de uso de sinónimos y de eliminación de palabras vacías produciría la siguiente consulta:

### Buscador

Cuál es el arma que más usan los caballeros Jedi?

### Buscador

Cuál es el arma <sup>favorita</sup> que más usan los caballeros Jedi?

### Buscador

Cuál es el arma <sup>favorita</sup> que más usan los caballeros Jedi? → Star Wars

### Buscador

~~Cuál es el~~ arma <sup>favorita</sup> que más usan los caballeros Jedi? → Star Wars



Buscador

arma favorita caballeros jedi star wars



Todo
 Imágenes
 Vídeos
 Noticias
 Shopping
 Más
 Herramientas

Aproximadamente 1.360.000 resultados (0,38 segundos)

<https://starwars.fandom.com/wiki/Jedi> > Leyendas**Jedi | Star Wars Wiki | Fandom**

Su arma era el sable de luz, un arma hecha de pura energía "un arma noble para otros tiempos mas civilizados". Los Jedi lucharon por la paz y la justicia en la ...

[https://starwars.fandom.com/wiki/Arma\\_de\\_la\\_Fuerza](https://starwars.fandom.com/wiki/Arma_de_la_Fuerza)**Arma de la Fuerza | Star Wars Wiki - Wookieepedia - Fandom**

La mayoría de los usuarios de este poder eran personas ajenas a la Orden Jedi ya que muchas de las culturas y tradiciones de la Fuerza se consideraron ...

<https://www.carteltec.com/star-wars-he-aqui-por-que-...>**Star Wars: he aquí por qué el sable de luz es el arma ...**

30 jul 2022 — El padawan luego cuestionará el uso de sables de luz por parte de los Jedi, creyendo que los cristales kyber podrían ser útiles de otra manera, ...  
Falta: eaballeros | Debe incluir lo siguiente: caballeros

<https://www.lacasadeel.net/curiosidades>**Star Wars: TOP 10 de los sables de luz más molones**

9 oct 2022 — El arma tiene un cristal kyber en su corazón que emite una hoja de plasma de color amarillo. Solo unos pocos Maestros Jedi han tenido uno, ...

<https://vandal.elespanol.com/.../Articulos>**Star Wars: Los 12 Jedi más poderosos de la saga - Vandal**

2 jun 2022 — Yoda, Luke Skywalker, Obi-Wan Kenobi, Mace Windu... Aquí repasamos cuáles son los maestros Jedi más fuertes que nos ha ofrecido La Guerra de ...



Otro factor a tener en cuenta es que, en general, es más aconsejable plantear los términos en singular que en plural. Más adelante en este curso, cuando se hable del denominado truncamiento esto se entenderá mejor. Además, la consulta puede incluir nombres propios y diferentes tipos de condiciones de búsqueda (tipo de documento, fecha de publicación, URL, entre otros). No obstante, si el sistema de recuperación de información que utilicemos tiene filtros de búsqueda, lo aconsejable es usarlos en lugar de emplear palabras clave para expresar esa condición. Por ejemplo, si buscamos resultados relativos a un determinado tipo de documento y existe ese filtro en el sistema, es mejor usarlo que incluir "artículo de revista" como palabra clave.

Acerca de si es más adecuado plantear la consulta en español o en inglés, si estamos ante un buscador web esta decisión dependerá del tipo de consulta del que se trate, mientras que si utilizamos un sistema de recuperación de información que cuente con índices, lo ideal es emplearlos y seleccionar el término que mejor se adecua a nuestra búsqueda. Además, como ya se ha indicado, en los sistemas que utilizan lenguaje documental para representar los conceptos de búsqueda de forma normalizada, los términos de nuestra consulta deberán estar expresados según lo determine el lenguaje documental (lista de encabezamientos de materias o tesauo) del sistema que estemos utilizando.



## REFERENCIAS BIBLIOGRÁFICAS

- **Blázquez Ochando, M.** (2015) Fuentes de información especializadas: Aspectos prácticos y teóricos.  
<http://mblazquez.es/wp-content/uploads/ebook-mbo-fuentes-especializadas.pdf>
- **Cid Leal, P.; Perpinyà Morera, R.** (2013). “Cómo y dónde buscar fuentes de información”. Barcelona: Universitat Autònoma de Barcelona.
- **Cordón García, José Antonio; Alonso Arévalo, Julio; Gómez Díaz, Raquel; García Rodríguez, Araceli.** (2016). “Las nuevas fuentes de información: la búsqueda informativa, documental y de investigación en el ámbito digital”. Madrid: Pirámide.
- **Martínez Rodríguez, Luis Javier** “Cómo buscar y usar información científica: Guía para estudiantes universitarios 2016”. (Unpublished).  
[http://eprints.rclis.org/29934/7/Como\\_buscar\\_usar\\_informacion\\_2016.pdf](http://eprints.rclis.org/29934/7/Como_buscar_usar_informacion_2016.pdf)