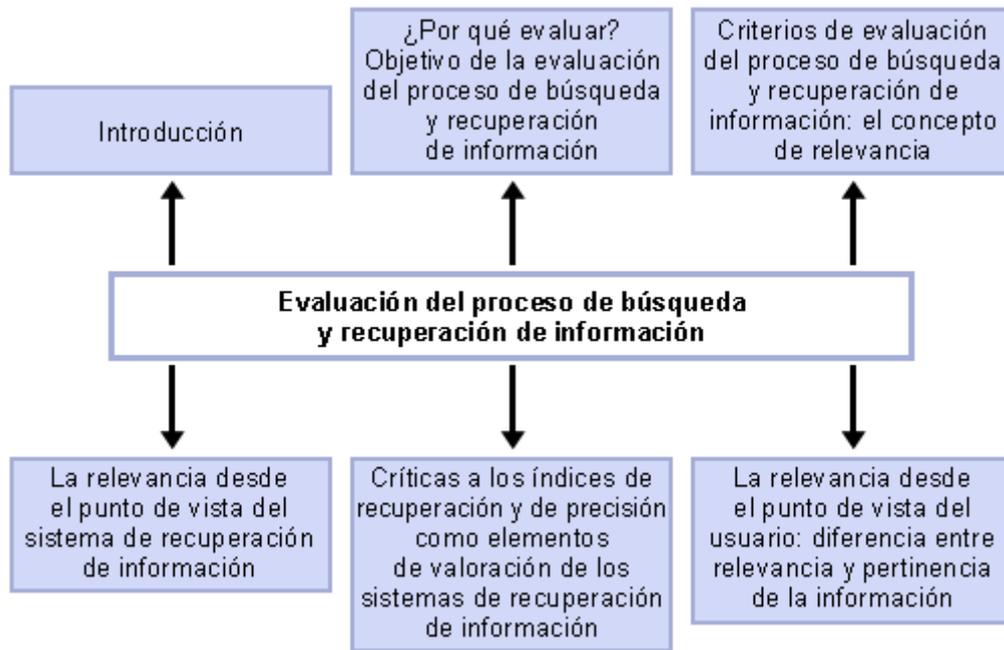


Evaluación del proceso de búsqueda y recuperación de información



Introducción

Antes hemos tratado las diferentes etapas del proceso de búsqueda y recuperación de información dando pautas de los procesos siguientes:

- Reconocer el problema que genera una necesidad de información que nos llega en forma de consulta.
- Seleccionar y valorar las fuentes de información que resultarán más adecuadas para resolver la consulta.
- Formular una estrategia de búsqueda que sea la traducción al lenguaje documental de la consulta que nos hace llegar el usuario.
- Ejecutar una estrategia en un sistema de recuperación de información (base de datos, catálogos de bibliotecas o un buscador de Internet).

Una vez que hemos pasado por todas estas fases, todavía tenemos que enfrentarnos con la última etapa de este proceso: la evaluación.

En este apartado haremos una breve introducción a la evaluación del proceso de búsqueda y recuperación de información, señalando, por una parte, cuál es su importancia en el proceso de búsqueda y, por la otra, cuáles son los principales criterios que se utilizan para ello.

¿Por qué evaluar? Objetivo de la evaluación del proceso de búsqueda y recuperación de información

A. Large, L.A. Tedd y R.J. Hartley (1999) dan respuesta a la pregunta con la que se inicia este apartado.

Con la evaluación del proceso de búsqueda y recuperación de información intentamos dar respuesta a las preguntas siguientes:

Evaluación del proceso de búsqueda y recuperación de información

<ul style="list-style-type: none">■ Preguntas que hacen referencia al sistema de recuperación de información o a la fuente de información.	<ul style="list-style-type: none">■ ¿Cuál es la mejor manera de diseñar un sistema de recuperación de información?■ ¿Qué ventajas y qué desventajas presenta el lenguaje controlado respecto del lenguaje natural en un sistema de recuperación de información?■ ¿Qué características tiene que reunir una interfaz a fin de que sea amigable para el usuario?
<ul style="list-style-type: none">■ Preguntas que hacen referencia a la técnica del proceso de búsqueda y recuperación de información.	<ul style="list-style-type: none">■ ¿Cuál es la estrategia de búsqueda más adecuada para cada consulta?■ Ante una consulta determinada, ¿qué estrategia nos permitirá recuperar toda la información que necesita el usuario sin recuperar también la que no necesita?
<ul style="list-style-type: none">■ Preguntas que hacen referencia al factor humano que interviene en el proceso de búsqueda y recuperación de información.	<ul style="list-style-type: none">■ ¿Cuáles son las características que debe reunir el documentalista ideal para resolver con éxito el proceso de búsqueda y recuperación de información?■ ¿Con qué criterios evalúa el usuario los resultados obtenidos en el proceso de búsqueda y recuperación de información?

Así, cuando hablamos de evaluación del proceso de búsqueda nos referimos a la evaluación de tres elementos inherentes a este proceso:

- Por una parte, las características del sistema de recuperación de información (la base de datos, el catálogo de biblioteca o el buscador de Internet) o la fuente de información que utilizamos con el fin de obtener la información que resolverá la consulta.
- Por otra parte, el modo en el que planteamos la búsqueda, que nos permitirá recuperar la información que resolverá la consulta.
- Finalmente, evaluamos al propio evaluador; es decir, evaluamos al documentalista (y sus criterios), ya que es quien determina en un primer momento hasta qué punto el sistema, la estrategia y los resultados obtenidos pueden resolver o no la consulta del usuario. También evaluamos al usuario, que es quien determina finalmente si los resultados cubren su necesidad de información.

Podemos evaluar por separado estos tres factores (sistema o fuente, técnica y factor humano) antes de iniciar el proceso de búsqueda y recuperación de información. De hecho, muchas de las pautas que hemos dado hasta ahora pueden servir como criterios de evaluación de los tres elementos.

Aquí, sin embargo, estamos hablando de una evaluación *a posteriori*; es decir, la que podemos llevar a cabo a la luz de los resultados obtenidos después de haber realizado todo el proceso (después de analizar la consulta, elegir la fuente de información que consideramos más adecuada y elaborar la estrategia de búsqueda que identificaba los términos de la consulta según el lenguaje propio de la fuente de información).

Cuando hablamos de evaluación del proceso de búsqueda y recuperación hacemos referencia a la evaluación que se hace a partir de los resultados de nuestras búsquedas. Mediante esta evaluación, podemos determinar lo siguiente:

- Si hemos analizado bien la consulta.
- Si hemos planteado bien la búsqueda.
- Si hemos elegido las fuentes de información oportunas.

Además, a partir de esta evaluación de los resultados podemos extraer conclusiones sobre los puntos siguientes:

- Los requisitos y las características que debe reunir un sistema de recuperación de información.
- Las habilidades y los conocimientos que debe reunir el documentalista con el fin de llevar a cabo este tipo de proceso.

Criterios de evaluación del proceso de búsqueda y recuperación de información: el concepto de relevancia

Después de plantear una búsqueda, lo primero que hace el documentalista es comprobar, a partir de los resultados, si la información que ha recuperado cumple las condiciones siguientes:

- Se corresponde con la estrategia de búsqueda que ha planteado.
- Sirve para solucionar la demanda de información del usuario, tanto por su calidad como por su cantidad.

En definitiva, el documentalista intenta determinar la relevancia de los resultados obtenidos.

No es conveniente definir el concepto de relevancia, básicamente por el hecho de que, tal y como señala F.W. Lancaster, se trata de una medida cualitativa subjetiva que puede ser valorada desde los dos puntos de vista siguientes:

- Desde el punto de vista del sistema: es la valoración que hace el documentalista de los resultados obtenidos, analizando si el sistema que ha utilizado, tal y como lo ha utilizado, le ha permitido recuperar la información que realmente pedía.
- Desde el punto de vista del usuario: es la valoración que hace la persona que pide la información de los resultados que le proporcionamos, determinando si realmente le sirven para solucionar su necesidad de información.

En este sentido, Alfons Cornella diferencia entre la *relevancia formal* (la que se valora desde el punto de vista del sistema) y la *relevancia semántica* (la que se valora desde el punto de vista del usuario).



Para acabar de entender estos conceptos, os aconsejamos que leáis el artículo de Alfons Cornella "La importancia de la *relevancia* en información", publicado en *Extra-net. Revista de infonomía: La información en las organizaciones* (Mensaje 305, 19/01/98):

<http://www.infonomia.com/extranet/index.asp?idm=1&idrev=1&num=305>

Cuando hablamos de relevancia estamos hablando de la relación que se establece entre la necesidad de información de un sujeto y la información que recupera con el objetivo de cubrir esta necesidad. Esto equivale a decir que estamos hablando de la esencia del proceso de búsqueda y recuperación de información.

La relevancia desde el punto de vista del sistema de recuperación de información

Cuando consultamos una base de datos o un buscador de Internet, obtenemos unos resultados en respuesta a una estrategia de búsqueda que hemos ejecutado previamente.

En este apartado podemos hacernos dos preguntas básicas:

- ¿Hasta qué punto hemos conseguido recuperar todos los documentos que existen en el sistema y que son relevantes para nuestra búsqueda?
- ¿Hasta qué punto hemos conseguido recuperar la máxima cantidad de documentos relevantes, sin recuperar los que no lo son?

Encontramos la respuesta a estas dos preguntas en dos índices básicos denominados:

- Índice de recuperación (o *recall ratio*).
- Índice de precisión (o *precision ratio*).

Estos dos indicadores fueron elaborados en el marco de los proyectos llevados a cabo a finales de los años cincuenta en el Cranfield College of Aeronautics de Inglaterra. El objetivo de estos proyectos era evaluar el efecto y las diferencias que el uso de vocabularios controlados y lenguajes naturales producían en el proceso de búsqueda y recuperación de información. Desarrollaron un estándar de evaluación, basado en dos medidas básicas: el *recall ratio* y el *precision ratio*.

- El índice de recuperación mide la proporción de documentos relevantes recuperados, es decir, mide la efectividad y exhaustividad de la búsqueda realizada, o hasta qué punto después de interrogar el sistema hemos conseguido recuperar todos los documentos relevantes para nuestra consulta.

Este índice se expresa en términos de proporcionalidad mediante la fórmula siguiente:

$$\text{Índice de recuperación} = \frac{\text{Número de ítems relevantes recuperados} \cdot 100}{\text{Número de ítems relevantes existentes en el sistema}}$$

- El índice de precisión mide la exactitud de la búsqueda, o hasta qué punto se ha conseguido recuperar sólo los registros que debían ser recuperados, evitando hacerlo con los que no eran necesarios.

Este índice se expresa en términos de proporcionalidad mediante la siguiente fórmula:

$$\text{Índice de precisión} = \frac{\text{Número de ítems relevantes recuperados} \cdot 100}{\text{Número de ítems totales recuperados}}$$

El índice de recuperación se relaciona con el concepto *silencio*, mientras que el índice de precisión se relaciona con el concepto *ruido*.

Hablamos de *silencio* en una búsqueda cuando no hemos conseguido recuperar todos los documentos pertinentes; o lo que es lo mismo, cuando hemos obtenido menos ítems de los que podíamos obtener.

Hablamos de *ruido* en una búsqueda cuando no hemos conseguido eliminar los documentos no pertinentes (cuando hemos obtenido más ítems de los que deberíamos haber obtenido).

De entrada, tanto el *silencio* como el *ruido* pueden ser causados por los hechos siguientes:

- Que no hayamos planteado correctamente la búsqueda.
- Que no hayamos elegido la fuente de información más apropiada.

La relación entre el índice de recuperación y el índice de precisión es expresada por varios autores mediante la matriz siguiente:

	Ítems recuperados	Ítems no recuperados
Relevantes	A	C
No relevantes	B	D

- **Grupo A:** engloba los ítems recuperados y relevantes para nuestra consulta.
- **Grupo B:** engloba los ítems recuperados y no relevantes para nuestra consulta.
- **Grupo C:** engloba los ítems que, a pesar de ser relevantes para nuestra consulta, no han sido recuperados.
- **Grupo D:** engloba los ítems que no han sido recuperados, pero que tampoco eran pertinentes para nuestra consulta.

A partir de la interpretación de esta matriz, podemos extraer las conclusiones siguientes:

- Podemos definir como búsqueda perfecta aquella en la que recuperamos todos los documentos relevantes sin ningún documento no relevante (es decir, sólo estarían llenos los grupos A y D, mientras que B y C se mantendrían vacíos).



Supongamos que partimos de una base de datos con 100 registros, de los cuales 40 son relevantes para nuestra consulta y los otros 60 no lo son.

A partir de diferentes resultados de nuestras búsquedas, medid los índices de recuperación y de precisión:

- Búsqueda A: recuperados 77 registros, de los cuales 28 son relevantes para la consulta.
- Búsqueda B: recuperados 40 registros, de los cuales 35 son relevantes para la consulta.
- Búsqueda C: recuperados 25 registros, de los cuales 10 son relevantes para la consulta.

-
- Los ítems que se engloban en el grupo B producen ruido en el resultado de la consulta, es decir, producen una reducción de la precisión de la búsqueda.
 - Los ítems que se engloban en el grupo C producen silencio en el resultado de la consulta, es decir, producen una reducción de el índice de recuperación de la búsqueda.

Además de la elaboración de estas dos medidas de evaluación, otra aportación importante del test de Cranfield fue la demostración de la relación inversa que existe entre los dos índices.

Para entender esta relación inversa, pondremos un ejemplo: imaginemos que queremos hacer una búsqueda que nos permita conseguir un índice elevado de precisión; esto implica recuperar necesariamente menos ítems (es decir, ser menos exhaustivos) mediante una estrategia de búsqueda muy esmerada.

En principio, lo ideal es que esta reducción de ítems se produzca por el lado de los ítems no relevantes (grupo D de la matriz), pero incluso podemos acabar eliminando también ítems relevantes que pasarían a engrosar el grupo C de la matriz, y eso repercutiría negativamente en el resultado del índice de recuperación.

Por el contrario, una elevada exhaustividad implica ampliar el número de ítems recuperados, tanto en el grupo A como el B de la matriz, por lo cual nuestra precisión bajaría.

▲ Índice de precisión ⇒ ▼ Índice de recuperación

▼ Índice de precisión ⇒ ▲ Índice de recuperación

Críticas a los índices de recuperación y de precisión como elementos de valoración de los sistemas de recuperación de información

Aunque ningún autor niega la valía de estos dos índices, presentan ciertas limitaciones como instrumentos de evaluación del proceso de búsqueda y recuperación de información.

Encontramos la primera de ellas, y también la más evidente, en el índice de recuperación; hablar de hasta qué punto hemos podido recuperar todos los documentos relevantes que se encuentran en el sistema implica partir de la premisa de que conocemos cuántos documentos relevantes hay.

Si hablamos de un sistema controlado y conocemos todos los documentos que lo constituyen, entonces sí podemos evaluar de manera exacta cuántos documentos relevantes no hemos sido capaces de recuperar.

Sin embargo, e indudablemente, esto no es posible cuando nos enfrentamos a sistemas cuya cantidad de información no permite determinar cuántos de los ítems que incluyen pueden ser relevantes para nuestra consulta. Sin duda, el mejor ejemplo actual de esta imposibilidad es Internet.

Por otra parte, también se argumenta en contra de estos indicadores que, a pesar de que pueden ser representativos de las características de un proceso de búsqueda y recuperación de información, no son los únicos elementos que sirven para su valoración final. Aspectos como el tiempo y el coste que debemos invertir en la consulta de un sistema de recuperación de información son algunos de los elementos que también es necesario tener en cuenta en el momento de hacer una evaluación. (¿Qué valoración puede hacerse de una base de datos que permita ser muy exhaustivo mediante unas estrategias de búsqueda altamente complejas?)

Otro tipo de críticas a estos índices hace referencia a su importancia relativa en el momento de valorar un sistema de recuperación de información. En este sentido, se trata de responder a la pregunta siguiente: para un usuario, ¿qué es más importante: un gran nivel de exhaustividad o un gran nivel de precisión?

La respuesta sería que depende del usuario, o dicho de otro modo, que depende de la necesidad de información del usuario. Para algunos usuarios es mejor un alto nivel de precisión, aunque esto implique sacrificar la exhaustividad (no es necesaria toda la información sobre el tema, basta con una proporción muy pequeña de la existente).

En este caso, el usuario valorará positivamente un sistema de recuperación que le permita obtener un resultado bajo, pero preciso, sobre lo que busca. En cambio, en el caso, por ejemplo, de un usuario que necesite conocer qué se ha publicado en los últimos años sobre un tema para hacer una tesis, se valorará más un resultado muy exhaustivo, aunque la precisión no sea muy elevada. Para su tranquilidad, en este caso el usuario preferirá asegurarse de que dispone de toda o casi toda la información.

Eso nos lleva al otro gran punto de vista de evaluación del proceso de búsqueda y recuperación de información: la valoración de la relevancia desde el punto de vista del usuario.

La relevancia desde el punto de vista del usuario: diferencia entre relevancia y pertinencia de la información

Como hemos visto anteriormente, el concepto de relevancia es una medida cualitativa altamente subjetiva. Esta afirmación todavía es más acertada cuando hablamos de relevancia desde el punto de vista del usuario (cómo mide un usuario la relevancia de los resultados obtenidos).

Para un usuario, la respuesta que recibe a su consulta (es decir, la información que le proporcionamos como respuesta a su consulta) es relevante en la medida en que le permitirá resolver su necesidad de información.

Por lo tanto, el criterio de valoración que utiliza el usuario es su necesidad de información. Sin embargo, ¿qué sucedería si esta necesidad de información no se hubiera reflejado bien en su consulta?

Recordad que definíamos la consulta como la expresión de la necesidad de información de un usuario, pero también señalábamos que muchas veces la consulta que nos hace llegar no refleja fielmente lo que necesita en realidad (en ocasiones porque no tiene claro qué busca y el documentalista no ha podido o no ha sabido ayudarlo a definir correctamente su necesidad de información).

En estos casos, los resultados podrían ser relevantes desde el punto de vista de la consulta. Responderían a lo que el usuario ha pedido, pero no serían pertinentes para su necesidad de información precisamente porque la consulta no refleja esta necesidad.

Recordad que:

- Desde el punto de vista del usuario, la relevancia es la correspondencia entre la información obtenida y su petición de información que nos ha hecho llegar.
- Desde el punto de vista del usuario, la pertinencia es la correspondencia entre la información obtenida y su necesidad de información.
- En este sentido, un resultado puede ser relevante pero no pertinente si la consulta no refleja la necesidad de información real del usuario.

Cuando intentemos resolver búsquedas temáticas, muy pocas veces conseguiremos que los primeros resultados que presentamos a nuestros usuarios sean valorados como totalmente pertinentes para su necesidad por el hecho de que la información obtenida les generará nuevas dudas y nuevas consultas.

En este sentido, aunque consigamos definir con bastante exactitud la necesidad de información del usuario, es posible que a la luz de unos primeros resultados se dé cuenta de que hay ciertos aspectos o matices que quizá le interesarían más, y que, por lo tanto, pida rehacer su consulta con el fin de darles cabida, o bien que, al recibir una primera respuesta, el usuario se dé cuenta de hasta qué punto la consulta que ha hecho llegar a un servicio no es representativa de lo que estaba buscando.

Por lo tanto, en este tipo de consultas es aconsejable que en el momento de presentar los primeros resultados aprovechemos para hablar con nuestro usuario sobre la forma en que se debe perfilar su consulta y llegar de manera pertinente a su resolución.

Otro problema que presenta la medida de la relevancia de los resultados de una búsqueda lo encontramos precisamente en el momento de esta medición.

En este sentido, A. Large y otros (1999) nos plantean la cuestión siguiente: ¿cómo se mide la relevancia de los resultados de una búsqueda: documento por documento o en conjunto?

Imaginemos que en respuesta a una consulta sobre las últimas investigaciones que se han hecho para encontrar una vacuna contra el sida presentamos al usuario (un estudiante de medicina que está elaborando un trabajo de final de carrera) un listado con 82 referencias de artículos y estudios publicados, que hemos extraído de la base de datos Medline.

En un primer momento, al usuario puede parecerle que, en general, los artículos encontrados son relevantes para su consulta, pero a medida que los vaya estudiando con más detenimiento empezará a hacer valoraciones del tipo "este documento es más o menos relevante que este otro".

Sin embargo, ¿puede ser una información medio relevante para una consulta? En principio, la respuesta a esta pregunta es no. La información o sirve al usuario para aclarar su necesidad de información o no le sirve.

Con esta cuestión logramos hacer más patente la dificultad de "medir" la relevancia de la información obtenida en un proceso de búsqueda, ya que en el fondo los criterios que utiliza el usuario son subjetivos.

Estos criterios subjetivos, tal y como señala F.W. Lancaster, dependen de los hechos siguientes:

- El estado cognitivo del sujeto.
- El problema que haya generado su necesidad de información y consulta posterior.
- Los conocimientos previos que tenga sobre el tema de su consulta.
- La urgencia que tenga para aplicar los conocimientos encontrados.
- El valor que otorgue a la información que le proporcionamos.

Por lo tanto, además de medir la relevancia, existen muchas variables individuales que se deben tener presentes, de las cuales destacamos las siguientes:

- a. Las características propias e individuales de cada usuario.



Para acabar de profundizar en este aspecto de la medida de la relevancia, os aconsejamos que leáis el siguiente artículo en el cual se describen los primeros resultados de un proyecto que tiene como objetivo determinar los principales criterios de evaluación de la relevancia utilizados por los usuarios de un servicio de información.

Schamber, Linda; Bateman, Judy (1996). "User Criteria in Relevance Evaluation : Toward Development of a Measurement Scale" [en línea]. *ASIS 1996, Annual Conference Proceedings* (octubre).

<http://www.asis.org/annual-96/ElectronicProceedings/schamber.html>

- b. Las preferencias del usuario ante diferentes alternativas.
- c. Los conocimientos previos del usuario de la materia tratada.

Un mismo ítem recuperado para una misma consulta realizada por distintos usuarios puede tener una relevancia diferente.

Ejercicios de autoevaluación

1. ¿A qué nos referimos cuando hablamos de evaluación del proceso de búsqueda y recuperación de información?
2. ¿Cuál es el objetivo de la evaluación del proceso de búsqueda y recuperación de información?
3. ¿Cómo podemos definir el concepto de relevancia desde el punto de vista del sistema de recuperación de información?
4. ¿Cómo podemos definir el concepto de relevancia desde el punto de vista del usuario?
5. ¿Qué mide el índice de recuperación?
6. ¿Cuál es el principal problema del índice de recuperación?
7. ¿Qué mide el índice de precisión?
8. ¿Por qué hablamos de relación inversa entre el índice de recuperación y el índice de precisión?
9. ¿Por qué los documentos no relevantes recuperados producen "ruido" en el resultado de una búsqueda?
10. ¿Qué diferencia hay entre relevancia y pertinencia desde el punto de vista del usuario?

Bibliografía

Cornella, A. (1998). "La importància de la 'rellevància' en informació" (en línea, mensaje 305, enero). *Extra-net. Revista de infonomia: La información en las organizaciones*.
<http://www.infonomia.com> (consulta: 24 de julio de 1999).

Lancaster, F.W.; Warner, A.J. (1993). "Relevance". *Information Retrieval Today* (pág. 54-61). Arlington: Information Resources Press.

Large, A.; Tedd, A.L.; Hartley, R.J. (1999). "Search Evaluation". *Information Seeking in the Online Age: Principles and Practice* (pág. 277-296). Londres: Bowker-Saur.

Schamber, L.; Bateman, J. (1996). "User Criteria in Relevance Evaluation: Toward Development of a Measurement Scale". *ASIS, Annual Conference Proceedings* (en línea, octubre).
<http://www.asis.org/annual-96/ElectronicProceedings/schamber.html> (consulta: 27 de julio de 1999).

Van Rijsbergen, C.J. (1979). *Information Retrieval*. "Chapter Six: Probabilistic Retrieval" (en línea). Londres: Butterworths. <http://www.dcs.gla.ac.uk/Keith/Chapter.6/Ch.6.html> (consulta: 20 de julio de 1999).

Van Rijsbergen, C.J. (1979). *Information Retrieval*. "Chapter Seven: Evaluation" (en línea). Londres: Butterworths. <http://www.dcs.gla.ac.uk/Keith/Chapter.7/Ch.7.html> (consulta: 20 de julio de 1999).