

> [Número 1, 2003](#) > [Evaluación RI](#)

La evaluación en recuperación de la información

Autora: Raquel Gómez Díaz (Universidad de Salamanca)

Citación recomendada: Raquel Gómez Díaz. *La evaluación en recuperación de la información* [en línea]. "Hipertext.net", núm. 1, 2003. <<http://www.hipertext.net>>

Sumario

1. [Resumen](#)
2. [Introducción](#)
3. [Antecedentes de la evaluación en r.i](#)
4. [La relevancia](#)
 - 4.1. [El cálculo de la relevancia](#)
5. [Principales medidas de evaluación en r.i.](#)
6. [La precisión](#)
7. [La exhaustividad](#)
8. [Relación entre la precisión exhaustividad](#)
9. [Medidas complementarias para la precisión y la exhaustividad](#)
 - 9.1. [Complemento del ratio de precisión](#)
 - 9.2. [Complemento del ratio de exhaustividad](#)
 - 9.3. [El índice de irrelevancia](#)
 - 9.4. [Complemento del índice de irrelevancia](#)
 - 9.5. [Generalidad](#)
 - 9.6. [La medida de f](#)
10. [La longitud de búsqueda esperada](#)
11. [Medidas relacionadas con el usuario](#)
12. [Bibliografía](#)
13. [Bibliografía complementaria](#)
14. [Notas](#)

1. Resumen

En este artículo se realiza una revisión de las medidas de evaluación que habitualmente se aplican a los sistemas de recuperación de la información siguiendo tanto el enfoque tradicional como aquel que evalúa la satisfacción de los usuarios, para los sistemas of line.

2. Introducción

Cuando se produce una necesidad informativa, mediante una estrategia de búsqueda más o menos complicada, interrogamos al conjunto de documentos, con el fin de obtener una respuesta que satisfaga la demanda. Para saber en qué medida la respuesta es satisfactoria, es necesario evaluar los resultados. Desde este punto de vista, la evaluación es la etapa final de la creación de un sistema.

La importancia de la evaluación en R.I. está muy ligada a la fase de investigación ya que sin unas medidas eficaces y estandarizadas, y colecciones experimentales adecuadas para este fin, no podremos hacer evaluaciones, ni lo que es más importante, no podremos comparar los sistemas de un modo fiable.

Dentro de la evaluación hay dos enfoques, el tradicional o algorítmico que trata de medir objetivamente como la respuesta es adecuada a la pregunta que hemos realizado, y el orientado a los usuarios, donde se

trata de medir la satisfacción del que ha hecho la demanda. Ambos enfoques no son excluyentes sino que son perfectamente complementarios [Ingersen 92]

En los últimos años, y debido a la importancia que tiene internet en la búsqueda y recuperación de documentos se están creando medidas específicas para evaluar estas recuperaciones como son la amigabilidad de los interfaces, la velocidad de la respuesta, los formatos de presentación, las conexiones con otros documentos... [Cacheda 01][Martínez 00] A continuación trataremos de los aspectos más importantes a evaluar desde el punto de vista algorítmico y de los usuarios, en los sitios de línea [1]

3. Antecedentes de la evaluación en r.i

El antecedente de los experimentos de evaluación está en el trabajo realizado por la ASTIA (Armed Services Technical Information Agency) y el College of Aeronautics sobre la recuperación de documentos representados con unitérminos extraídos del título y el resumen [Ellis 90]. En este experimento realizado en 1953 fue donde se utilizó por primera vez el concepto de relevancia, aunque éste ya había sido formulado en la década anterior [Saracevic 75].

El primer trabajo de evaluación propiamente dicho, fue el desarrollado en Crandfield [Cleverdon 66] a finales de la década de los 50. La importancia de estos trabajos radica en que fueron los primeros en establecer la metodología de la evaluación y las herramientas que debían emplearse. Las herramientas son: Una colección de documentos de la que se extraen las preguntas y los juicios de relevancia. Con éstos se calculan las medidas de precisión y exhaustividad para analizar los resultados y establecer las comparaciones entre los modelos. Esta metodología es la que sigue presente en la evaluación de la recuperación [Harter 97].

4. La relevancia

Uno de los principales problemas en R.I. es la variedad de interpretaciones de algunos conceptos, como es el caso del de relevancia [Mizzaro 98].

Es importante definir este concepto, porque está en la base del resto de las medidas que tradicionalmente se vienen aplicando en R.I.. Aunque se formuló entre los años 30-40 no se utilizó experimentalmente hasta el test de Crandfield.

El concepto de relevancia se ha estudiado desde distintos puntos de vista [Saracevic 97] : lógica, filosofía, psicología, semántica, documentación... Estos enfoques los podemos resumir en dos tendencias: la relevancia objetiva y la subjetiva. La primera hace hincapié en los sistemas, normalmente define cómo la materia de la información recuperada coincide con la de la pregunta. La subjetiva, es la que tiene en cuenta al usuario [Swanson 86]. Dentro de este enfoque está la relevancia mirada desde el punto de vista del usuario [Schamber 90], [Wilson 73]. Para Schamber la relevancia se refiere a la utilidad, o potencial uso de los materiales recuperados, con relación a la satisfacción de los objetivos, el interés, el trabajo o los problemas intrínsecos del usuario.

En la relevancia subjetiva, se estudia desde el punto de vista de la información nueva que consigue un usuario de un documento. Según este concepto, la información conocida no es relevante [Boyce 92]. Hay autores a caballo entre estas dos tendencias, para los que la relevancia tiene un componente objetivo y otro subjetivo. Así Barry [Barry 94], determina la relevancia de un documento en función de siete criterios (1. Información que contiene un documento; 2 experiencia previa del usuario; 3 creencias y preferencias del usuario; 4 otras informaciones y fuentes; 5 fuentes del documento; 6 documento como entidad física; 7 situación de los usuarios) de los cuales dos son objetivos (1 y 5) y cinco subjetivos (2, 3, 4, 6 y 7).

Harter [Harter 96] indica que el principal problema de los estudios sobre los factores que afectan a la relevancia es que se han hecho de manera intuitiva, tal vez esto sea debido la variedad de interpretaciones de este término.

Muy ligado al concepto de relevancia está el de pertinencia; con frecuencia se entremezclan y confunden. Según Korfhage [Korfhage 97], relevancia es la medida de cómo una pregunta se ajusta a un documento,

(esta visión coincide con el enfoque de la relevancia objetiva) y pertinencia es la medida de cómo un documento se ajusta a una necesidad informativa (lo que otros autores definen como relevancia subjetiva).

Es decir, según este autor, la diferencia entre uno y otro radica en cómo expresamos la necesidad de información, por lo tanto, a la hora de establecer la relevancia tenemos que tener en cuenta la doble dificultad que lleva implícita la pregunta, porque tiene que ser el reflejo de la necesidad informativa (de ella dependerá la pertinencia) y al mismo tiempo tiene que ser adecuada para la búsqueda de los documentos que resuelvan la necesidad informativa, ya que la relevancia va a depender directamente de la formulación concreta de la demanda informativa. A pesar de que Korfhage establece esta distinción entre relevancia (relevancia objetiva) y pertinencia (relevancia subjetiva), no todos los autores siguen esta línea, sino que algunos los utilizan como sinónimos.

En el caso de los trabajos en español muchas veces se han traducido los dos términos indistintamente para referirse a los dos conceptos. La valoración de la pertinencia es mucho más difícil de realizar ya que es el propio usuario el único que sabe si un documento se ajusta a su necesidad o no. Además la pertinencia en un mismo usuario cambia de un momento a otro, ya que la información conocida no es pertinente, puesto que no resuelve la necesidad informativa.

Para calcular la relevancia, lo más habitual es establecer valores binarios: un documento es relevante, es decir, sirve como respuesta a nuestra pregunta, (valor 1) o no sirve (valor 0), aunque también se puede fijar una gradación, y establecer una escala ordinal para medir la relevancia de los documentos [Cuadra 67]. El problema de determinar una escala es que no hay una guía clara para elaborarla. Por ejemplo Keen [Keen 71], usa cuatro valores de escala, para dividir del más relevante al menos relevante. Saracevic [Saracevic] [88] da tres valores a su escala: relevante, parcialmente relevante y no relevante, pero en la práctica distinguir entre un documento relevante y uno parcialmente relevante es muy difícil.

4.1. El cálculo de la relevancia

Existen dos métodos para calcular la relevancia, uno manual y otro conocido como polling :

Manual: consiste en la exploración de los documentos uno a uno para saber si se adecúan o no como respuesta a una pregunta. Muchas veces establecer la relevancia de un documento para una pregunta determinada resulta difícil y los especialistas no se ponen de acuerdo, por ello, es conveniente que los juicios los haga más de uno, y a ser posible un número impar de especialistas. El principal problema que presenta este método, es que en colecciones muy grandes, hay que invertir gran cantidad de tiempo, lo que supone mucho dinero para realizar esta operación y esto no siempre es posible. Además, algunas bases de datos son más especializadas que otras, lo que hace necesario contar con un número mayor o menor de especialistas. Para solventar estos problemas se crean las colecciones experimentales, donde se fija de antemano qué documentos son relevantes para cada pregunta.

Estas colecciones suelen tener un tamaño medio y suelen pertenecer a una misma área temática o muy próxima para que no sea necesaria la intervención de muchos especialistas. Un ejemplo de una colección manual es la de Crandfield [Cleverdon 91]. En este caso se buscaron los artículos y se les pidió a los autores que elaboraran preguntas cuya respuesta fuera su artículo y también se les pidió que citaran otros artículos que correspondieran a esa misma pregunta que ellos habían formulado. Con las preguntas y los artículos citados por los autores se elaboró la base de datos, la colección de preguntas, y los juicios de relevancia.

Polling : cuando las bases de datos son muy grandes, y no es posible evaluar uno a uno los documentos, para determinar cuáles son los documentos relevantes, se recurre al "polling". Lo que se hace es analizar de manera manual un número determinado de documentos recuperados con distintos sistemas, este número suele ser elevado (varios centenares) y se corresponde con los primeros documentos recuperados con cada sistema. Este conjunto de documentos es el que de manera manual analizan los expertos, que son los encargados de decir en último término si son relevantes o no. Este sistema asume que la gran mayoría de los documentos relevantes son encontrados, si no por todos los sistemas, sí al menos por alguno de ellos, y los no recuperados pueden considerarse como no relevantes [Kowalski 97].

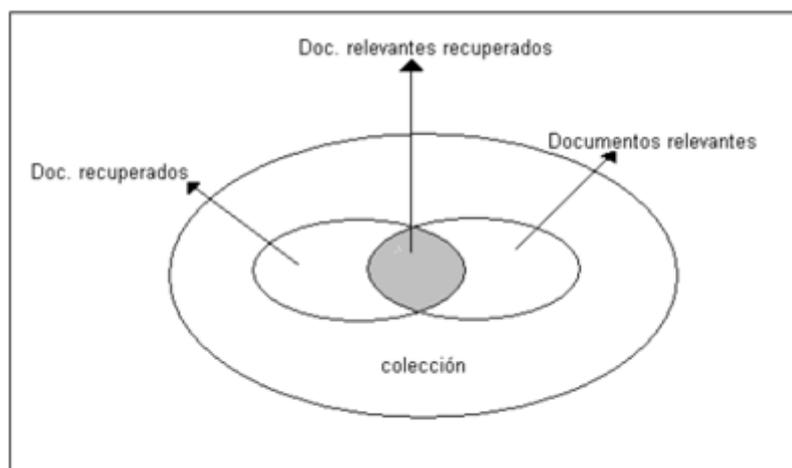
De esta manera no es necesario evaluar toda la base de datos, pero aun así el sistema es fiable ya que el número de documentos que se suele examinar es elevado. Este sistema es el que se viene utilizando en las TREC desde 1994 [Harman 95].

5. Principales medidas de evaluación en r.i.

Una vez definido el concepto de relevancia y relacionando éste con si un documento es recuperado o no, podemos establecer una serie de medidas que nos servirán para evaluar los sistemas de recuperación. A continuación expondremos las principales medidas comunes a todos los modelos de recuperación.

Los documentos pueden ser recuperados o rechazados al establecer la comparación entre la pregunta y la base de datos. El conjunto de documentos recuperados se divide, salvo en los sistemas perfectos, en dos grupos: documentos relevantes recuperados, es decir aquellos que se han recuperados correctamente y los no relevantes, recuperados erróneamente que provocan ruido en la salida. Los documentos no recuperados, que a su vez se dividen en los relevantes, rechazados por el sistema de manera errónea y los no relevantes, rechazados de manera correcta por el sistema. Esto mismo lo podemos ver en el siguiente dibujo.

Ilustración 1 Esquema recuperación documentos. Fuente:[Baeza-Yates 1999]



6. La precisión

Este concepto fue definido por Kent [Kent 55], como factor de pertinencia. Hay otros autores que se refieren a él, como ratio de aceptación. Para Salton [Salton 83], la precisión es la proporción de material recuperado realmente relevante, del total de los documentos recuperados. A esta definición Frakes [Frakes 92] añade que el resultado de esta operación está entre 0 y 1. Así, la recuperación perfecta es en la que únicamente se recuperan los documentos relevantes y por lo tanto tiene un valor de 1.

En esta medida, se evalúa directamente la correlación de la pregunta con la base de datos e indirectamente sirve para ver cómo es de completo el algoritmo de indización [Kowalski 97]. Si el algoritmo de indización tiende a generalizar teniendo un umbral alto en los términos de índice o al usar los conceptos genéricos de indización, entonces la precisión es baja, no importa cómo sea el algoritmo de similitud entre la pregunta y el índice.

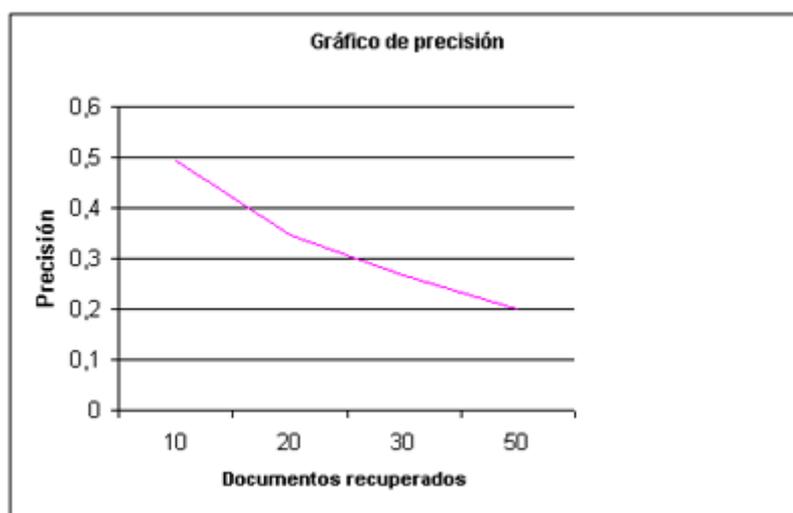
Ecuación 1 Precisión. Salton

$$\text{Precisión} = \frac{\text{Documentos relevantes recuperados}}{\text{Documentos recuperados}}$$

Esta medida está relacionada con dos conceptos, el de ruido y el de silencio informativo. De este modo, cuanto más se acerque el valor de la precisión a 0, mayor será el número de documentos recuperados que no le sirvan al usuario y por lo tanto el ruido que encontrará será mayor.

La salida obtenida en la recuperación es ordenada en función de la relevancia, por lo que los documentos más relevantes están al comienzo de la salida, de esta manera a medida que avanzamos en el número de documentos recuperados, la precisión decae.

Su representación gráfica se hace marcando en el eje de las x el número de documentos y en las de las y, los valores de precisión de 0 a 1, asociada a esos documentos recuperados de modo que los sistemas más precisos son aquellos que en su gráfica describen una curva con valores altos al principio y que van decreciendo. Comparando las distintas curvas de los sistemas, podemos hacernos una idea clara de cuáles son más precisos.



7. La exhaustividad

La exhaustividad, aunque en menor medida que la precisión es el otro concepto más utilizado en la evaluación de los sistemas de recuperación.

Muchos autores, por influencia del término inglés la denominan "recall" o "rellamada". Es la proporción de material relevante recuperado, del total de los documentos que son relevantes en la base de datos, independientemente de que éstos, se recuperen o no. Esta medida es inversamente proporcional a la precisión. Fue formulada, al igual que la de precisión por Kent [Kent 55], con el nombre de factor de exhaustividad. Años más tarde, Swet#n28 [Swet 63] la llamó probabilidad condicional de un ítem, y Goffman y Newil [Goffman 64] la denominaron sensibilidad (sensitivity).

La ecuación propuesta por Salton [Salton 83]:

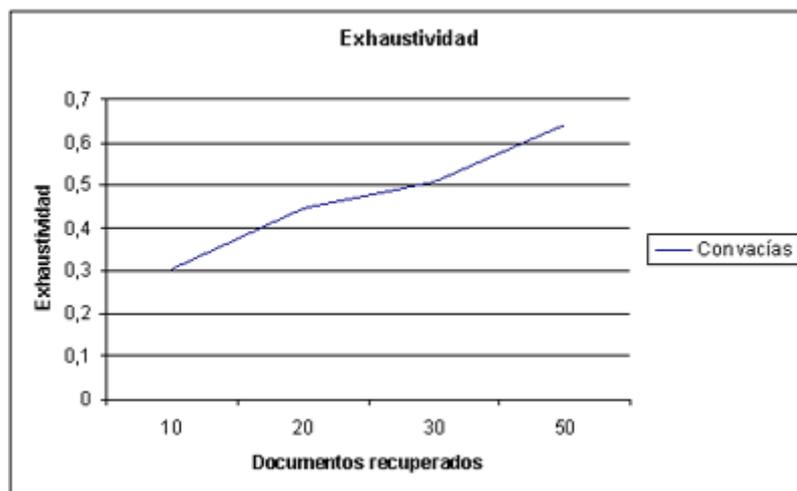
Ecuación 2 Exhaustividad. Salton

$$\text{Exhaustividad} = \frac{\text{Documentos relevantes recuperados}}{\text{Documentos relevantes}}$$

Si el resultado de este cálculo tiene como valor 1, tendremos la exhaustividad máxima, ya que hemos encontrado todo lo relevante que había en la base de datos, por lo tanto no tendremos ni ruido ni silencio informativo: la recuperación será perfecta.

Para alcanzar una exhaustividad alta, es necesario utilizar como índice términos generales de alta frecuencia, es decir, que aparezcan en muchos documentos de la colección. Para alcanzar una precisión alta, es necesario que los términos aparezcan con frecuencia alta, pero en pocos documentos y con nula en el resto. Aunque para el usuario la situación ideal es una precisión y exhaustividad alta, lo que Cooper denomina utilidad teórica, [Cooper 76] y esto es imposible.

Al igual que la precisión también podemos representarla gráficamente, para ello en el eje de las x marcamos el número de documentos y en el de las y el valor de la exhaustividad calculada para cada documento. A medida que aumenta el número de documentos recuperados, recordemos que la salida es ordenada en función de la relevancia, la exhaustividad va en aumento. El comportamiento normal de esta gráfica, es que la curva vaya aumentando. Los sistemas serán más exhaustivos cuando alcancen al principio valores altos (próximos a 1), y después vayan disminuyendo.

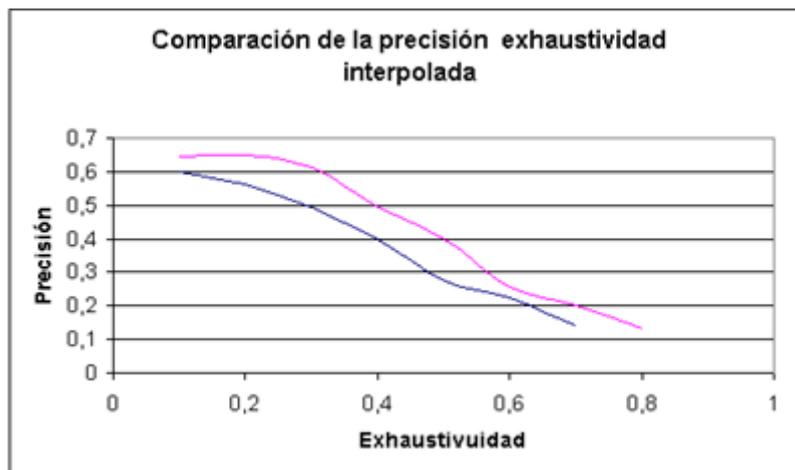


Korfhage señala principalmente dos objeciones a los sistemas que se basan en la precisión y en la exhaustividad. El primero de ellos es que mientras que la precisión se puede determinar, la exhaustividad no, ya que para calcularla necesitamos previamente el número de documentos relevantes, precisamente para evitar esto se utiliza el polling. El segundo de los puntos que señala Korfhage es que la exhaustividad y la precisión son igualmente significativas para los usuarios. Mientras que unos prefieren una precisión mayor, otros prefieren una exhaustividad más alta, (incluso esto varía en función del tipo de necesidad informativa) y ambas cosas es imposible tenerlas al mismo tiempo.

8. Relación entre la precisión exhaustividad

Necesitamos comprobar que la precisión y la exhaustividad están compensadas, ya que un sistema con una exhaustividad muy alta pero con baja precisión y viceversa no será adecuado. Para comprobar como se relacionan la precisión y la exhaustividad en una sola gráfica, podemos hacerlo de varias maneras: calculando la precisión exhaustividad interpolada: es decir tomamos un conjunto de documentos y calculamos para cada valor de precisión su exhaustividad. Por ejemplo tomamos los veinte primeros documentos recuperados, donde hay quince documentos relevantes y calculamos la precisión y la exhaustividad para cada documento recuperado (si el primer documento recuperado es relevante tendremos una precisión de 1/1 y una exhaustividad de 1/15). También podemos hacerlo de manera no interpolada, en este caso calculamos la exhaustividad por tramos de documentos recuperados. Por ejemplo tomamos veinte documentos y calculamos el valor de exhaustividad en los cinco primeros documentos recuperados, luego en los diez, luego en los quince y finalmente en los veinte documentos recuperados.

Una vez que tenemos estos valores, en ambos casos marcamos los puntos, en el eje de las x los valores correspondientes a la exhaustividad y para cada valor de ésta marcamos en el de las y el valor de la precisión que le corresponde. Uniendo los puntos obtenemos la curva que nos dice cómo se relacionan estas dos medidas en cada sistema y comparándolas ver qué sistema es el más efectivo.



En 1983 Salton y MacGill, sugirieron un método para la evaluación del sistema proponiendo salidas ordenadas de los documentos en las respuestas. De este modo, la precisión y la exhaustividad dependían del valor de corte, es decir, del punto a partir del cual se considera que al usuario ya no le interesan los documentos. Este criterio Blair lo denomina " punto de futilidad " [Blair 80]. La precisión y la exhaustividad se calcula para cada posición en la lista de documentos recuperados.

9. Medidas complementarias para la precisión y la exhaustividad

Existen otra serie de medidas complementarias a la precisión y a la exhaustividad, la mayor parte mucho menos utilizadas que éstas.

9.1. Complemento del ratio de precisión

También se le denomina " factor de ruido ". Consiste en los documentos no relevantes recuperados partido por los recuperados.

Ecuación 3 Complemento del ratio de precisión

$$\text{Complemento del ratio de precisión} = \frac{\text{Documentos no relevantes recuperados}}{\text{Documentos recuperados}}$$

9.2. Complemento del ratio de exhaustividad

Su ecuación se calcula dividiendo los documentos relevantes no recuperados entre el total de los documentos relevantes.

El primero en formularlo fue Swets 1963 [Swets 63] que lo denominó probabilidad condicional de una pérdida. En 1964 Fairthorne Fairthorne 64] lo denominó ratio del esnobismo. ("snobbery ratio")

Ecuación 4 Complemento del ratio de exhaustividad

$$\text{Complemento del ratio de precisión} = \frac{\text{Documentos no relevantes recuperados}}{\text{Documentos recuperados}}$$

9.3. El índice de irrelevancia

Este índice se obtiene de dividir los documentos recuperados no relevantes a la pregunta entre el total de los documentos contenidos en la colección. Como muchas de las medidas anteriores fue formulada en primer lugar por Swets en 1963, que se refirió a él como probabilidad condicional de bajada falsa (conditional probability of false drop). Cleverdom, Mills and Keen [Cleverdon 66] la llamaron posteriormente fallout. También ha sido denominada "desechado" (discard).

Ecuación 5 Índice de irrelevancia

$$\text{El índice de irrelevancia} = \frac{\text{Documentos no relevantes no recuperados}}{\text{Documentos no relevantes}}$$

Según Kowalski [Kowalski 97] con esta medida podemos establecer con qué efectividad está actuando un sistema de recuperación. Esta medida es el inverso de la exhaustividad y nunca nos encontraremos con un resultado de 0/0, a menos que todos los documentos sean relevantes para la búsqueda.

9.4. Complemento del índice de irrelevancia

Swets en 1963, lo denominó "probabilidad condicional de una correcta respuesta negativa" (conditional probability of a correct rejection). Goffman and Newill la llamaron "especificidad". Se calcula dividiendo los documentos no relevantes no recuperados entre el total de los documentos no relevantes:

Ecuación 6 Complemento del índice de irrelevancia

$$\text{Complemento del índice de irrelevancia} = \frac{\text{Documentos relevantes}}{\text{Total de documentos}}$$

Con las siguientes medidas podemos poner en relación las medidas anteriores.

9.5. Generalidad

La generalidad sirve para calcular la densidad de documentos relevantes [Korfage 97]. Se calcula dividiendo los documentos relevantes entre el total de los documentos de la base.

Ecuación 7 Generalidad

$$\text{Generalidad} = \frac{\text{Documentos recuperados}}{\text{Número de documentos}}$$

La precisión, la exhaustividad, el índice de irrelevancia y la generalidad se relacionan mediante la siguiente ecuación:

Ecuación 8 Relación entre precisión, exhaustividad, y generalidad

$$\frac{P}{Ir} = \frac{P/(1-P)}{G/(1-G)}$$

Donde $P/(1-P)$ es el ratio de los documentos relevantes recuperados partido el de los no relevantes recuperados. $G/(1-G)$ es el ratio de los documentos relevantes en la colección partido los documentos no relevantes en la colección.

P/r es la ejecución de la recuperación en los documentos relevantes entre la ejecución de la recuperación en los documentos no relevantes. Es deseable tener el primero de los dos alto.

9.6. La medida de f

Sirve para corregir el error de la Distancia, en los casos en los que la exhaustividad (E) y la precisión (P) se compensan. Su ecuación es:

Ecuación 9 Medida de F

$$F_{\beta} = \frac{(B^2 + 1) \cdot P \cdot R}{B^2 \cdot P + R}$$

Donde B es un valor preestablecido, teniendo en cuenta que si B es igual a uno, estamos dando la misma importancia a P que a E, si B mayor que uno de damos más importancia a E y si es menor de damos más importancia a P.

10. La longitud de búsqueda esperada

Es el número de documentos no buscados que el usuario puede esperar examinar antes de encontrar el número de documentos deseados [Cooper 68].

11. Medidas relacionadas con el usuario

La precisión y la exhaustividad se basan en que el conjunto de documentos recuperados para una preguntas es el mismo, independientemente del usuario. Sin embargo, lo habitual es que la valoración de la respuesta obtenida, varíe de unos usuarios a otros, o incluso en un mismo usuario dependiendo del momento de la recuperación, por este motivo son necesarias las medidas orientadas a los usuarios ya que ellos son la razón de ser de la existencia del sistema. La efectividad de un sistema es una medida ajena al propio sistema que relaciona la satisfacción del usuario con la salida que el sistema proporciona. Medir la satisfacción del usuario resulta muy importante, pero es complicado y es menos objetivo que las medidas vistas anteriormente, por eso estas medidas se han ido dejando de lado.

Veamos el siguiente esquema. Ilustración 2 Recuperación. Medias orientadas a los usuarios [Fuente Baeza-Yates 99]

En función de este esquema podemos definir las siguientes medidas

Ratio de cobertura : es la proporción de documentos relevantes conocidos por el usuario que son actualmente recuperados.

Ratio de novedad : proporción de documentos relevantes recuperados que previamente son conocidos por el usuario

Exhaustividad relativa : ratio de documentos relevantes recuperados, examinados por el usuario, partido por el número de documentos que el usuario quiere examinar.

Supongamos que el usuario conoce 15 documentos relevantes, y el sistema recupera 10 relevantes, incluyendo 4 documentos que son conocidos por el usuario. El ratio de cobertura sería $4/15$ es decir 26,6%. De aquí el usuario puede inferir que hay aproximadamente 38 documentos relevantes, aproximadamente cuatro veces el número de documentos recuperados. Si el usuario ha visto 6 nuevos documentos relevantes

añadidos a esos 15 previamente conocidos, podemos estimar que la base de datos contiene 16 ó 17 documentos relevantes que él nunca ha visto y a partir de aquí puede intentar recuperarlos, modificando, si lo considera oportuno, su estrategia de búsqueda.

Siguiendo con el ejemplo, el ratio de novedad sería 6/10. Un ratio de cobertura alto, podría dar al usuario alguna confianza en que los sistemas localicen todos los documentos relevantes. También sugiere que el sistema es efectivo en la localización de documentos desconocidos para el usuario. Del ejemplo anterior, el usuario puede inferir que aproximadamente el 60% de algún grupo de documentos relevantes recuperados para esta pregunta y esta base de datos, en particular, no será previamente conocida. Por supuesto, al usuario no le interesa saber que puede recuperar, aquellos documentos que él ya conoce, por lo tanto, es deseable que el ratio de novedad sea alto. En cuanto a la exhaustividad relativa, puede referirse más directamente a la cuestión de cómo el usuario quiere algunos documentos. Supongamos que el sistema presenta 20 documentos al usuario y que éste quiere 5 documentos relevantes.

Si solo hay 3 documentos relevantes entre los 20, la exhaustividad relativa será 3/5, el usuario solo obtiene 3 de los 5 que busca. Si por el contrario, hay 5 o más documentos relevantes entre los 20, entonces, presumiblemente el usuario podrá abandonar después de encontrar los 5 deseados con una exhaustividad relativa de 5/5 es decir de 1. Si la exhaustividad relativa es de 1, la medida falla al referirse los esfuerzos a localizar los documentos.

Podría ser que el usuario encuentre los documentos entre los primeros 5 ó 6 examinados o podría ser que necesitara examinar los 20, por lo tanto esto nos da pie para definir una nueva medida: esfuerzo de exhaustividad, que es el ratio del número de documentos relevantes deseados partido por el número de documentos examinados para encontrar el número de documentos relevantes deseados. Esta medida asume que la colección contiene el número de documentos relevantes deseado y que el sistema de recuperación permite al usuario localizarlos todos, lo cual aunque es deseable no siempre es posible. Este ratio puede ir de 1, si los documentos relevantes deseados son los primeros documentos examinados por él, a próximo a 0, si el usuario necesita examinar un gran número de documentos para encontrar los pocos que desea.

Otras medidas relacionadas con el usuario son la utilidad y satisfacción. De las medidas vistas hasta ahora, éstas son las más subjetivas, por lo que habrá que valorarlas con mucho cuidado. La satisfacción pone énfasis en la coincidencia entre lo que el usuario quiere y lo que el usuario recibe.

12. Bibliografía

- [Baeza-Yates 99] Baeza-Yates, Ricardo, Ribero-Neto, Berthier. Modern Information Retrieval. New York: Addison-Wesley, 1999
- [Barry 94] Barry, C.L. User -defined Relevance Criteria: An Exploratory Study Journal of the American Society for Information Science 1994 45 (3) p. 149-159
- [Blair 80] Blair Searching bases in large interactive document retrieval systems Journal of the American Society for Information Science 1980 (31) 4 p. 271-277
- [Boyce 92] Boyce, B. Beyond Topically: A two storage view of relevance and retrieval process Information procesing and Management 1992 18 p. 105-109
- [Cacheda 01] Cacheda, Fidel, Viña, Ángel Simulación para la evaluación de sistemas de recuperación en el WWW. [Consultado el 1-10-2002
- [Cleverdon 66] Cleverdon C. W. , J. Mills and Keen, E. M. ASLIB Cranfield proyect: Factors Determining the perfomance indexing Systems. 1966
- [Cleverdon 91] Cleverdon, C.W.. The Significance of The Cranfield Tests on Index Languages. In A. Bookstein, editor, Proceedings of the 14 th Annual International ACM/SIGIR Conference on Research and Development in Information Retrieval, Chicaco, Illinois, USA, October 1991.
- [Cooper 68] Cooper W. S. Expected Search Lenght: A single measures of retrieval effectiveness based on the weak ordering action retrieval systems. American Documentation 1968; 19 p. 30-41
- [Cooper 76] Cooper, S. The Paradoxical Role of Unesamined Documents in the Evaluation of Retrieval Effectiveness. Information Processing and Management 1976 12 p. 367-375
- [Cuadra 67] Cuadra, A. C. And Katter, R. V. Opening the blok box of "relevance ". Journal of documentation 1967 23 (4) p 291-303

[Ellis 90] New horizons in Information Retrieval. London: Library Association, 1990

<[Fairthorne 6] Fairthorne, R. A. Basic parameters or retrieval text. Proceedings of 1964 annual Meeting of the American Documentation Institute Washington: Spartan Books, 1964 pp 343-347

[Frakes 92] Frakes, W. B. and Baeza Yates, R. (ed.) Information Retrieval: data structures and Algorithms. Mexico: Prentice-Hall, 1992

[Goffman 64] Goffman and Newill Methodology for test and evaluation of information retrieval systems. Information Storage and Retrieval (1964) 3 p. 19-25

[Harman 95] Harman, D. Overview of the Third Text Retrieval Conference (TREC-3) [en línea] <trec.nist.gov/pubs/trec3/t3_proceedings.html > [consultado el 07/10/02]

[Harter 96] Harter, P. Variations in Relevance Assessment and Measurement of Retrieval Effectiveness. Journal of the American Society for Information Science 47 (1) 1996 p. 37-49

[Harter 97] Harter, S.P. Hert, C. A. [1997] Evaluation of Information Retrieval Systems: Approaches, Issues, and Methods Annual Review of Information Science and Technology 1997 (ARIST) vol 32, pp. 3-94

[Ingersen 92] Ingersen, P. Information Retrieval interaction. London: Taylor Graham, 1992.

[Keen 71] Keen, E. M. Evaluation Parameters, in G. Salton (Editor): The SMART Retrieval System, Prentice Hall, Englewood Cliffs, 1971.

[Kent 55] Kent A. Et al. Machine literature searching. VIII. Operational Criteria for Designing Information Retrieval Systems American Documentation Abril 1955 6 (2) p. 93-101

[Korfhage 97] Korfhage, R., Information Storage and Retrieval, New York.: John Wiley, 1997.

[Kowalski 97] Kowalski, G. Information Retrieval Systems Theory and Implementation. Boston: Kluwer Academic Publishers, 1997

[Martinez 00] Martínez Méndez, Javier. Aproximación general a la evaluación de la recuperación de la información por medio de los motores de búsqueda en Internet . [Consultado el 1-10-2002

<http://www.um.es/qtiweb/fjmm/ibersid2000.PDF>]

[Mizzaro 98] Mizzaro, Stefano How many relevances in information retrieval? Department of Mathematics and Computer Science University of Udine Via delle Scienze, 206 --- Loc. Rizzi --- 33100 Udine --- Italy

[Salton 83] Salton, G. y M. J. McGill.. Introduction to Modern Information Retrieval. New York: McGraw Hill. 1983

[Saracevic 75] Saracevic, T.. Relevance: A review of and a framework for the thinking on the notion in information science. Journal of the American Society for Information Science, 26(6):321- 343, 1975.

[Saracevic 88] Saracevic, T. [et al.] A study of information seeking and retrieving, background and methodology. Journal of the American Society for Information Science , 39 (3) p. 161-176

[Saracevic 97] Saracevic, T.. Relevance: A review of and a framework for the thinking on the notion in information science. (paper review) En Readings in Information Science edited by Karen Spark Jones, Peter Willet. San Francisco: Morgan Kaufmann Publisher 1997.

[Schamberg 90] Schamberg, L., Einseberg, B. and Nilo, S. A re-examination of relevance: toward a dynamic, situational definition Information Processing and Management, 1990, (6) p. 755-775.

[Swanson 86] D. R. Wanson Subjective versus objective relevance in bibliographic retrieval system. Library Quarterly 1986 56, p. 389-398

[Saracevic 75] Saracevic, T.. Relevance: A review of and a framework for the thinking on the notion in information science. Journal of the American Society for Information Science, 26(6):321- 343, 1975.

[Saracevic 88] Saracevic, T. [et al.] A study of information seeking and retrieving, background and methodology. Journal of the American Society for Information Science , 39 (3) p. 161-176

[Saracevic 97] Saracevic, T.. Relevance: A review of and a framework for the thinking on the notion in information science. (paper review) En Readings in Information Science edited by Karen Spark Jones, Peter Willet. San Francisco: Morgan Kaufmann Publisher 1997.

[Schamberg 90] Schamberg, L., Einseberg, B. and Nilo, S. A re-examination of relevance: toward a dynamic, situational definition Information Processing and Management, 1990, (6) p. 755-775.

[Swanson 86] D. R. Wanson Subjective versus objective relevance in bibliographic retrieval system. Library Quarterly 1986 56, p. 389-398

[Swets 63] Swets, J. A. Information retrieval Systems Science, 141 (3577): July 1963 p. 245-250

[Wilson 73] Wilson, P. Situational relevance Information Storage and Retrieval 1973 9 p. 457-469

13. Bibliografia complementaria

Bates, M. Bate's bibliography <http://www.gseis.ucla.edu/faculty/bates/bib-intro.html>
Borlum, P. and Ingwersten, P. [1997]. The development of method for evaluation of interactive Information Retrieval System. Journal of Documentation 1997 53 (3) p. 225-250
Hert, C. A. Understanding information retrieval interaction: theoretical and practical implications. Greenwich: Ablex Publishing Corporation, 1997.
Keen, E.M. [1971] Evaluation parameters. En SALTON, G. (ed), The SMART retrieval system Experimentes in automatic document processing. New Jersey:Prentice-Hall, 1971 p 74-111
Keen, E.M. [1996] Measures and Averaging Methods Used in Performance Testing Indexing System. Crandfield, Eng.,Aslib Crandfield Project 1966

Lancaster, W. F. And Warner, A. J. [1993] Information Retrieval Today. Arlington: Information Resources Press, 1993
Schamberg, L. [1994] Relevance and Information behaviours. Annual Review of Information Science and Technology (ARIST) 1994 29 p. 3-48
Su, L. T. [1994] The Relevance of Recall and Precision in User Evaluation. 1994 Journal of the American Society for Information Science 1994 45 (3) p. 207-217
Voorhers, Ellen Philosophy of I.R. Evaluation [consultado 23-10-02] <http://www.ercim.org/publication/ws-proceedings/CLEF2/vorhees.pdf>]
Yao, J. J. Measuring Effectiveness Bases on User Preference of Documents. Journal of the American Society for Information Science 46, (2), 1995 p. 133-145

14. Notas

[1] Para los sistemas on line. Walter, Geraldine amd Janes, Joseph. On line retrieval. Colorado: Libraries unlimited, 1993. [[volver](#)]



Last updated 02-08-2019
© Universitat Pompeu Fabra, Barcelona