

# Síntesis y crítica de las evaluaciones de la efectividad de los motores de búsqueda en la Web.

[Francisco Javier Martínez Méndez](#) y [José Vicente Rodríguez Muñoz](#)

Grupo de Tecnologías de la Información, Universidad de Murcia.

Campus Universitario de Espinardo s/n 30.071-Murcia (España).

## Resumen

La necesidad de un análisis crítico de la evaluación de los sistemas de recuperación de información subyace tras su traslado al contexto de la web. En este nuevo escenario han surgido nuevos tipos de sistemas y de problemas que han dado lugar, al desarrollo de una serie de trabajos específicamente concebidos para analizar la viabilidad de la recuperación de información. En estos trabajos destacan tres grandes grupos: los dedicados a analizar las características formales o externas del sistemas, los que someten a estos sistemas a algún tipo de ensayo o experimentación y finalmente, una serie de propuestas orientadas a proponer una metodología científica de carácter global para su evaluación.

## [Abstract in English](#)

## Introducción.

El paralelismo existente entre el crecimiento de la información depositada en diversas fuentes de la web y la proliferación del número de motores de búsqueda o directorios que aseguran poseer la mayor cantidad de recursos debidamente indexados y accesibles a través de ellos, ha sido un fenómeno imposible de evitar en tanto que, al poco tiempo de la aparición de estos ingenios y su popularización en el seno de los usuarios de la red de redes, se generó la siguiente incógnita: "¿cuál es el mejor de estos sistemas?". De hecho, casi todos los usuarios de Internet creen poseer la respuesta a esta pregunta, pero se debería reflexionar detenidamente antes de contestar, ¿cuáles son las bases sobre las que fundamentarían su respuesta?

El conjunto de argumentos a utilizar puede ser variado: familiaridad con un motor determinado, pleno convencimiento de su calidad, desconocimiento del resto, costumbre de uso de un sistema, etc. Sea cual fuere la respuesta, lo que está claro es que sobre estos nuevos sistemas de recuperación de información sigue latente una necesidad intrínseca de evaluación.

Con el fin de hallar una respuesta, ha proliferado la producción científica en este campo, así Oppenheim opina que estos trabajos "han ganado popularidad a medida que la web ha crecido" ([Oppenheim, 2000](#)), en tanto que el mayor número de recursos está repercutiendo en la calidad de la respuesta ofrecida por estos sistemas.

Los resultados que ofrecen estos estudios de evaluación muestran resultados dispares y dispersos. La *disparidad* proviene de la obtención de conclusiones diferentes entre la mayor parte de los estudios, aunque siempre se pueden identificar algunas reflexiones comunes en todos ellos. La *dispersión* surge cuando el conjunto de motores evaluados es funcionalmente distinto, lo que dificulta la especificación de conclusiones válidas y generales, al evaluar en cada estudio un conjunto de motores que poco tiene que ver con el evaluado por otro. Incide también en esta dispersión la no repetición de estos estudios en el tiempo, lo que conlleva a que los mismos ofrezcan conclusiones válidas sobre una situación determinada, al mismo tiempo que resultan incapaces de percibir la evolución de esta tecnología, aspecto verdaderamente preocupante si se consideran los continuos cambios que se producen en este campo, tanto de carácter técnico como de incremento de información.

También existen grandes diferencias entre los métodos de evaluación aplicados y en el alcance. De hecho, existe una considerable cantidad de trabajos donde la evaluación se limita a aspectos explícitos: externos, formales o

testimoniales, del motor de búsqueda (amigabilidad de la interfase, velocidad de respuesta, formatos de presentación, documentación existente, ayuda del sistema, etc.). Si bien estos aspectos poseen su importancia, no permiten interpretar, en modo alguno, la efectividad de las operaciones de recuperación de información.

Un segundo conjunto de trabajos se centra más profundamente en el estudio de esta efectividad, haciendo uso, en la mayoría los casos, de las medidas basadas en la realización de juicios de relevancia, exhaustividad y precisión generalmente ([Lancaster, 1973](#)). El uso de estas medidas, por sí solas, tampoco llega a ofrecer conclusiones definitivas y concluyentes sobre las prestaciones de los motores de búsqueda (sin olvidar la subjetividad que presentan ambas ratios), aunque sí proporcionan resultados mucho más sólidos y argumentados que las evaluaciones meramente formales.

A la hora de abordar el diseño de un estudio de este tipo, junto a los juicios de relevancia hay que tener presente la serie de problemas vinculados con la propia naturaleza de la web, contexto donde se presentan situaciones que en el entorno de los SRI tradicionales resultarían de difícil imaginación. Esto se debe, entre otras muchas razones, a que el propio origen de la web poco tiene que ver con la estructura de un SRI, sino más bien con la de un sistema orientado a la difusión de información y de trabajo en grupo. Oppenheim presenta varios de estos problemas: presencia de registros duplicados, tendencia a recuperar documentos poco relevantes, recuperación de enlaces a páginas web que ya no existen, el "spamming", alineación inadecuada de los documentos recuperados, etc. Con todo ello, para este autor "el problema fundamental de la web es su tamaño, heterogeneidad e inconsistencia; los recursos cambian y se multiplican y la impresionante velocidad de respuesta de los motores de búsqueda no está ligada, generalmente, a una efectiva recuperación de información" ([Oppenheim, 2000](#)).

Esta atípica naturaleza propicia que no puedan emplearse exclusivamente las medidas basadas en la relevancia, sino que sea necesario establecer un grupo de medidas específicas al nuevo contexto, tales como la ratio de enlaces fallidos, el grado de solapamiento, acierto único y cobertura del motor. El cálculo de estas medidas, unidas generalmente a la precisión y la exhaustividad, conforman un segundo grupo de estudios, que pueden denominarse como estudios experimentales, en tanto que suelen llevar a cabo algún ensayo antes de mostrar sus conclusiones.

Recientemente ha proliferado un tercer grupo de trabajos orientados hacia el establecimiento de un marco global para la evaluación de los motores de búsqueda, más que a un análisis comparativo de la efectividad de unos motores frente a otros. Aunque el objeto principal de estos trabajos varía considerablemente de uno a otro y, por lo tanto, resulta complicado establecer una línea común entre ellos, sí coinciden en su pretensión de no ofrecer como resultado qué motor es más preciso o cuál de ellos ha crecido más en el último trimestre, sino en la idea de concebir una propuesta integral de evaluación de estos sistemas.

## Estudios explícitos.

Las primeras evaluaciones explícitas de los motores de búsqueda datan del año 1995, aproximadamente un año después de la aparición de los primeros motores. A continuación se discutirán los parámetros más utilizados en estas evaluaciones, con referencia a los estudios que han contado con ellos en su realización y las conclusiones generales de cada uno de ellos.

### Aspectos formales y prestaciones en la recuperación de información.

Los primeros estudios suelen basarse, preferentemente, en las características formales del motor de búsqueda y en las descripciones técnicas del sistema. Chu y Rosenthal ([Chu, 1996](#)), mencionan a Courtois, Baer y Stark, quienes destacan la potencialidad del motor Webcrawler en todo lo relacionado con su flexibilidad a la hora de plantear las ecuaciones de búsqueda y su rápida respuesta, resaltando además, su interfase como muy adecuada para usuarios poco iniciados. Basándose también en la idea de la flexibilidad de la interfase, Scoville apuesta por los motores Excite, Infoseek y Lycos. Chu y Rosenthal citan también el trabajo de Kimmel, quien califica a Lycos como el mejor a partir del estudio de la documentación aportada y de sus características externas.

Davis considera el tamaño del índice del motor y las posibilidades de recuperación de información, decantándose por Alta Vista, Hot Bot e Infoseek, sobre un total de siete motores evaluados ([Davis, 1996](#)). Westera analiza las posibilidades que ofrece la interfase de usuario en la recuperación de información. La autora divide las capacidades de búsqueda en dos grupos: básicas y especiales, en el primero destacan Alta Vista y Google y en el segundo Alta Vista y Hot Bot ([Westera, 2001](#)).

Chu y Rosenthal comentan finalmente que C|net, empresa de evaluación de servicios de información en línea, comparó diecinueve motores de búsqueda, considerando el acierto, la facilidad de uso y la cantidad de opciones avanzadas que proporcionaban. Como resultado final destacaban a Alta Vista ([Chu, 1996](#))).

Slot analizó exhaustivamente, en 1996, los contenidos manejados por dieciséis motores de búsqueda y directorios, considerando también objeto de su estudio, entre otras variables, el tiempo de respuesta y la claridad de la interfase. En un segundo nivel de análisis, estudió en profundidad las posibilidades que cada motor ofrecía en la recuperación de información. En el primer apartado consideraba a Alta Vista y Yahoo como excelentes; en el segundo apartado seguía destacando Alta Vista ligeramente sobre los demás (Slot, 1996).

También en 1996, Zorn, Emanoil y Marshall comparan cuatro motores de búsqueda (Alta Vista, Infoseek, Lycos y Opentext), analizando el número de documentos estimado de sus índices, su documentación, la porción de página indexada, sus características de búsqueda, la presencia de documentos duplicados en el índice y, finalmente, si empleaban algún método de alineamiento en la presentación de los documentos ([Zorn, 1996](#)). Los resultados de este estudio mostraron a Alta Vista y Lycos como los motores de mayor tamaño, y mucha similitud en cuanto a las capacidades de búsqueda y formatos de presentación de los documentos. La documentación aportada por Alta Vista y Opentext se considera excelente y todos indexan la página completa a excepción de Lycos.

## Tamaño del índice del motor de búsqueda.

Idealmente, si un motor de búsqueda recopilara en su índice la totalidad de los documentos de la web (o un porcentaje cercano), sin duda alguna, ese motor sería el predilecto de todos los usuarios de Internet, otorgándose a este parámetro un valor prioritario por encima de otros. La realidad es bien distinta, "el constante cambio y expansión de la web, provoca que ninguno de los motores de búsqueda pueda indexar la totalidad de sus documentos. Muchos estudios, realizados para estimar el tamaño de la web, han determinado que los motores recopilan entre el 5% y el 30% de la totalidad de documentos de la web, y la unión de los once principales motores de búsqueda no alcanza el 50%" ([Chang, 2001](#)). Aún así, la estimación del tamaño de los índices de los motores y la determinación aproximada del porcentaje de web recopilada, ha sido objeto de interés de varios autores.

Uno de los trabajos más citados es el realizado por Lebedev en 1996, quien realizó ocho preguntas a distintos motores. Cada pregunta la formaba sólo una palabra clave, relacionada con la Química o con la Física, posteriormente procedió a la suma del total de los documentos devueltos por cada motor en cada búsqueda. Lebedev también analizó la incorporación de nuevas páginas al motor. Anteriormente se ha citado que Slot también analizó el tamaño del índice en su estudio, resaltando el inmenso tamaño de Alta Vista frente a los demás analizados. ([Slot, 1996](#)).

El tamaño del índice del motor, el número de documentos devueltos, características relacionadas con el modo de almacenar las referencias y el formato de salida de los documentos, constituyen el objeto del trabajo de Peterson. Este autor tomó datos en tres períodos de tiempo distintos (febrero, mayo y noviembre del año 1996), interrogando a ocho motores por medio de dos expresiones de búsqueda, una conformada por un término individual y la otra por una frase literal. El resultado de esta parte del experimento mostraba al motor Hot Bot como el de mayor número de documentos ([Peterson, 1997](#)). El número de documentos y las posibilidades de búsqueda también son objeto de interés para Maldonado y Fernández, quienes han hecho uso en esta estimación, de una utilidad presente en los motores Alta Vista e Infoseek que permite recuperar páginas que apuntan (enlazan) a una determinada dirección. El cómputo final de los resultados situaba a Yahoo (un directorio), como el sistema que más páginas tenía recogidas, seguido por los motores Alta Vista y Excite. El análisis de las posibilidades de búsqueda estableció grandes similitudes entre Alta Vista e Infoseek.

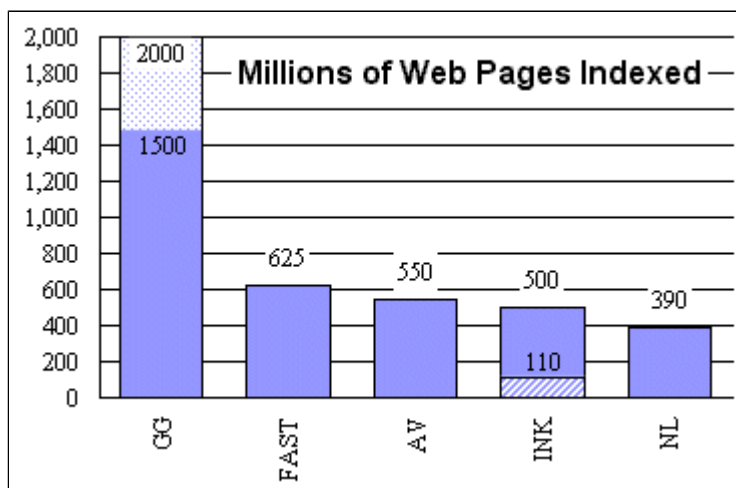
La obsolescencia que posee esta medida, implica la necesidad de actualizar los estudios con cierta frecuencia, tal como realiza Notess en la web Searchengine Showdown ([Notess, 2002a](#)), donde presenta, entre otra serie de datos, el número de documentos recuperados por distintos motores tras la realización de veinticinco consultas (Tabla 1):

Motor	Tamaño estimado	Extensión anunciada	% diferencia
Google	968	1.500	54,9
WISEnut	579	1.500	159

Alltheweb (Fast)	580	507	-13,5
Northern Light	417	358	-15,2
Altavista	397	500	25,9
Hotbot	332	500	50,6
MSN Search	292	500	71,23

**Tabla 1:** Clasificación de los motores según tamaño y determinación de diferencias entre tamaño estimado y declarado. Fuente: [Notess, 2002d](#).

La siguiente ilustración recoge la estimación del tamaño del índice de los motores de búsqueda, que ha realizado Sullivan en la web *Searchenginewatch*.



GG=Google, FAST=FAST, AV=AltaVista INK=Inktomi NL=Northern Light.

Ilustración 1 Clasificación de los motores de búsqueda según tamaño.

Fuente: [Sullivan, 2001a](#)

Este estudio sitúa a Google en primer lugar destacado y a Fast en segundo lugar. Estos resultados coinciden con el anterior de Notess (con la excepción del motor WISEnut). Google aparece con dos cifras distintas porque realmente "Google sólo ha indexado 1 billón y medio de páginas (1000 millones es 1 billón), pero gracias al uso extensivo de los enlaces que realiza este motor, puede actualmente devolvernos listas de páginas adicionales que nunca ha visitado" ([Sullivan, 2001a](#)). Otro caso particular es Inktomi, que diferencia entre dos conjuntos de páginas: "Lo mejor del web", que agrupa a 110 millones de documentos y el "resto del web", que agrupa 390 millones de páginas.

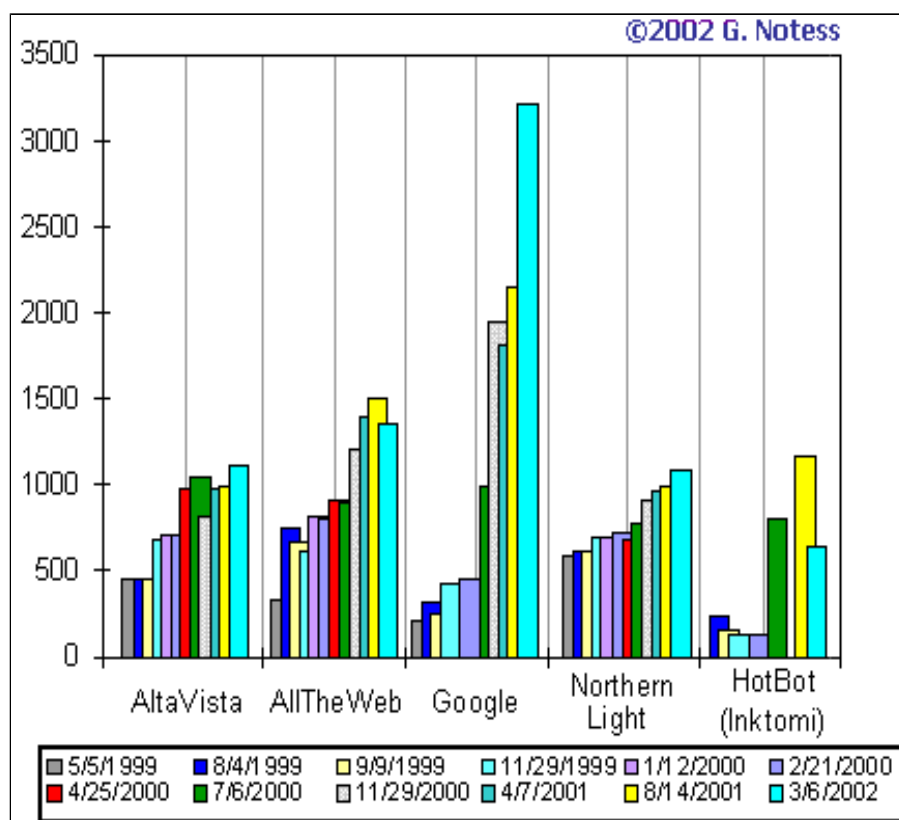


Ilustración 2 Evolución en el tiempo del tamaño de los motores de búsqueda.

Fuente: [Notess, 2002b](#)

Este informe de Sullivan analiza también la evolución del tamaño de índice a lo largo del tiempo (Ilustración 2). Alta Vista era el motor de mayor tamaño, siendo igualado a partir de junio de 2000 por Fast, siendo ambos motores rebasados en septiembre de 2000 por la vertiginosa ascensión de Google. Esta evolución es similar a la presentada por Notess en el trabajo anteriormente citado.

## Audiencia del motor.

Otro criterio explícito empleado en las evaluaciones de los motores de búsqueda es su nivel de audiencia, es decir, el número de accesos que computan sus sitios web.

En la web *Searchengineswatch.com* se recogen diversos trabajos de Sullivan donde se utiliza esta medida, sintetizando datos procedentes de empresas especializadas en cálculos de audiencia. El primero que presenta es el realizado por Jupiter Media Matrix, a partir de los accesos que más de cien mil usuarios de Internet realizan sobre los motores o directorios analizados. Este informe muestra como motor más popular a Yahoo (64.8%) seguido por Microsoft Network (58.3%) y American Online (46%) ([Sullivan, 2001b](#)). Si bien este estudio viene avalado por su interactividad con los usuarios y por su actualización, introduce una considerable distorsión al mezclar directorios con motores. No obstante, del mismo se desprende los tres motores anteriormente mencionados siguen una perceptible tendencia de crecimiento positivo de su audiencia. Otro dato interesante es la clara evolución positiva de Google y la negativa de Alta Vista, aspecto ya destacado por otros autores.

Sullivan analiza también el estudio confeccionado por [Nielsen/NetRatings](#), de actualización semanal donde se mide el porcentaje de usuarios que visitan los distintos sitios web junto al tiempo medio que le dedican, además de medir la audiencia según el número de accesos computados. No todos los sitios web que presenta este análisis son motores de búsqueda, incorporando también directorios y sitios web de otra naturaleza, tales como las web de Microsoft, Amazon o Disney, por ejemplo. La serie de datos ofrecidos sitúa a Yahoo en primer lugar (48%) y a Microsoft Network en segundo lugar (38,5%). El primer motor que encontramos es Google (16,2%), situado en tercer lugar, seguido de Lycos en cuarto lugar con algo más de un once por ciento ([Sullivan, 2001c](#)).

## Número de consultas realizadas.

En otro trabajo, Sullivan proporciona una clasificación del número de consultas que reciben cada día los motores,

## Tráfico redirigido.

**Porción de la página indexada.**

## Fidelidad.

### Sinopsis de estudios explícitos.

[illegible]



Maldonado	X		X						Alta Vista
Notess (a)			X						Google
Notess (b)				X					Google
Sullivan (a)			X						Google
Jupiter Media					X				MSN
Nielsen Ratings					X				MSN
Statmarket						X			Alta Vista
Sullivan (d)								X	Google
Winsip	X	X	X				X	X	Lycos
Stobart								X	Alta Vista

**Tabla 2:** Sinopsis de los 20 estudios explícitos analizados que evalúan los motores de búsqueda. PB: posibilidades de búsqueda. T: tiempo de respuesta. GI: Interface gráfica de usuario. DOC: documentación. TI: tamaño del índice. CRE: crecimiento porcentual del tamaño del índice ACI: acierto. FA: frecuencia de actualización. AUD: audiencia. TR: tráfico redirigido. PO: porción de página indexada. FO: formato de presentación y alineamiento. FID: fidelidad. NC: consultas recibidas por motor

El número medio de características empleadas por estudio es, a nuestro parecer, bastante bajo (1.9 por trabajo). Las características más utilizadas son: tamaño del índice (7 veces) , posibilidades de búsqueda (6 veces) e interfase de usuario (5 veces). Hasta siete motores aparecen citados como los mejores según el objeto del estudio.

Una agrupación con base en esta serie de características más estudiadas, permite extraer una serie de coincidencias:

- En los estudios que analizan el tamaño del índice, los resultados eran favorables al motor Alta Vista hasta que en los mismos se reflejó la presencia de Google, siendo este motor el que ocupa la posición de privilegio desde entonces.
- Cuando los estudios se basan en el análisis de las posibilidades de búsqueda, es Alta Vista el más veces destacado
- No existe una tendencia común de resultados en los estudios analizan la interfase gráfica de usuario.

## Estudios experimentales.

A continuación se presentan los estudios más significativos que, dentro de este campo, han sido desarrollados en los últimos cinco años. El propósito de esta revisión no es otro que mostrar cómo han evolucionado estos procesos de análisis y cómo se han ido incorporando medidas cada vez más complejas, que van adaptándose a las características del contexto en el que trabajamos, la web.

### Chu y Rosenthal (1996).

El punto de partida de esta revisión, no puede ser otro que el trabajo elaborado por Chu y Rosenthal para la Conferencia Anual de ASIS de 1996, quienes comentaban que las diferencias de opinión de los estudios explícitos residían básicamente en la ausencia de una metodología clara para la evaluación y se preguntaban "si podíamos llegar a pensar en desarrollar esa metodología que facilite unos resultados que ayuden al usuario a discernir qué herramienta es la más apropiada para sus necesidades" ([Chu, 1996](#)).

Los autores reutilizan las medidas propuestas por Cleverdon (cobertura, exhaustividad, precisión, tiempo de respuesta, esfuerzo del usuario y formato de presentación de los documentos), adaptadas al entorno de la web, considerando que siguen siendo válidas casi cuatro décadas después de enunciarse. Así, los autores plantean un método de evaluación donde se valoran cuatro criterios:

1. Composición de los índices: la Cobertura, la Frecuencia de actualización y la Porción de página indexada.

2. Capacidades de búsqueda: las prestaciones que los distintos motores poseen a la hora de recuperar información y la serie de operadores de búsqueda disponibles.
3. Ejecución de la recuperación de información: la Precisión, la Exhaustividad y el Tiempo de respuesta, procedentes todos ellos de las propuestas clásicas de Lancaster y Fayen ([Lancaster, 1973](#)).
4. Esfuerzo del usuario: la Documentación y la Interfase del sistema ([Chu, 1996](#)).

Esta propuesta constituye un substancial salto cualitativo frente a las anteriores evaluaciones explícitas, al mismo tiempo que se aprovecha de aquéllas los parámetros más fiables, se incorpora una serie de cuantificaciones que proporcionan una mejor ponderación de la efectividad de los SRI en la web. Otro aspecto a resaltar de este estudio es su fecha de realización, apenas dos años después de la popularización de los sitios web y casi contemporáneo al desarrollo de los primeros motores de búsqueda, lo que le confiere un carácter ciertamente embrionario dentro de una línea seguida a continuación por otros autores. El único aspecto negativo a señalar de este estudio es el escaso número de motores analizados: Alta Vista, Lycos y Excite, por lo que hace falta recurrir a otras experimentaciones para alcanzar una idea más completa sobre el estado de la cuestión, aunque quizá el escaso número de motores desarrollados en ese momento limitó el alcance del mismo.

### **Leighton y Srivastava (1995-1999).**

De similar importancia al trabajo anterior son las aportaciones de Leighton y Srivastava. En el primero de sus trabajos, Leighton evaluaba cuatro motores: Infoseek, Lycos, Webcrawler y WWWorm. Tras la formulación y ejecución de ocho preguntas, calculaba una serie de medidas basadas en la relevancia, considerando además la ratio de acierto único (documentos recuperados sólo en un motor), y el número de enlaces erróneos junto al número de documentos duplicados. El resultado final de este estudio destaca a Lycos e Infoseek sobre los otros dos. El impacto de este trabajo sorprendió hasta al mismo autor, quien al principio de la página web donde se recoge una copia de este artículo, indica textualmente: "me encuentro sorprendido con la popularidad de este estudio". Sorprendido porque los datos incluidos en el mismo tenían un corto período de vida y estoy seguro de que los resultados se encuentran absolutamente obsoletos" ([Leighton, 1995](#)).

Esta plena conciencia de obsolescencia le lleva a realizar otros trabajos en el mismo campo, contando ya con la colaboración de Srivastava. En el primero de ellos procede a la evaluación de cinco motores de búsqueda: Alta Vista, Excite, Hotbot, Infoseek, y Lycos. El período de evaluación abarcaba de enero a marzo de 1997 y en el mismo, los autores introdujeron una interesante variante a la típica manera de calcular la precisión, diseñando una función de evaluación que confiere un peso específico a la capacidad de colocar documentos relevantes dentro de los primeros veinte entregados como respuesta al usuario. Esta función, denominada "First 20 Precision" mide, al mismo tiempo, la precisión y el acierto de mostrar los documentos relevantes antes que los documentos que no son relevantes. En este estudio se penaliza la existencia de enlaces inactivos, de manera que aquellas páginas que no hayan sido actualizadas hace bastante tiempo influirán de forma decisiva en los resultados ([Leighton, 1997](#)). Los resultados obtenidos en el estudio de la precisión destacan a Alta Vista, Excite e Infoseek (por este orden). Cuando se considera el alineamiento, se aprecian pocas diferencias pero cambia el orden: Infoseek, Excite y Alta Vista ([Leighton, 1999](#)).

### **Arents (1995-2001).**

Algunos estudios explícitos han evolucionado en estudios experimentales en sus sucesivas repeticiones. Un ejemplo es el trabajo desarrollado por Arents desde 1995. Este autor selecciona distintos motores de búsqueda y los clasifica según una escala: {Mejor-Muy bueno-Bueno-Útil}, basando sus apreciaciones en la facilidad de uso y en la efectividad del motor, entendida ésta como la cantidad, precisión y legibilidad de los documentos recuperados.

### **Bharat y Broder (1997).**

Este estudio pretende medir "los tamaños relativos de los índices de los motores de búsqueda analizados y su grado de solapamiento" ([Bharat, 1998a](#)). Los motores analizados fueron cuatro: Alta Vista, Excite, Hot Bot e Infoseek, y se tomaron datos en dos períodos de tiempo, junio y noviembre de 1997. Con relación al solapamiento, los cambios producidos en este período de seis meses fueron insignificantes, estimándose del 1.4%. Los tamaños relativos estimados presentan cifras consistentes de un experimento a otro, aunque cambia el orden, ya que en junio Hot Bot supera a Alta Vista, Excite e Infoseek (por este orden) y en noviembre, Alta Vista supera a Hot Bot.



## **Ralph (1997).**

Ralph realizó veinticinco preguntas, la mayor parte de ellas correspondientes a términos individuales, nombres de personas o instituciones o materias, sobre siete motores y un directorio (Yahoo). Así analizó el número de documentos recuperados, la precisión, la exhaustividad relativa y el alineamiento. Para estudiar este último aspecto, valoraron la presencia de los documentos relevantes al principio de la secuencia de documentos entregados como resultado. Los resultados de la precisión obtenidos en este estudio dividen a los motores analizados en dos grupos claramente diferenciados. El primero de ellos agrupa a Alta Vista, Excite, Hot Bot e Infoseek (que alcanzan porcentajes superiores al 60%), frente al resto de motores cuyos valores oscilan entre el 30% de Lycos y el 35% de Webcrawler. El motor que mejor porcentaje alcanza de exhaustividad relativa es Hot Bot, que supera ligeramente el 25%. La ponderación del alineamiento de los documentos igualaba a Hot Bot y Excite.

## **Wishard (1998).**

Wishard determinó la precisión de diecisiete motores de búsqueda, realizando sobre ellos una serie de preguntas relacionadas con la Geología y no encuentra diferencias significativas entre todos ellos.

## **Lawrence y Giles (1998).**

La cobertura de los motores de búsqueda centra el principal interés de este estudio. Sus autores estimaron también el tamaño de los distintos índices, su grado de solapamiento y el porcentaje de enlaces erróneos que proporcionaba cada motor. El conjunto de seis motores evaluados era: Alta Vista, Excite, Hot Bot, Infoseek, Lycos y Northern Light.

Como resultado de este análisis, se deduce que el motor Hot Bot, con un 57.5%, es el motor que más documentos comunes posee en su relación, seguido de Alta Vista (con un 46.5%). En la estimación del tamaño de la web, se tasa en un 34% el porcentaje indexado por Hot Bot. La unión del índice de este motor y del índice de Alta Vista, produciría un conjunto de más de 320 millones de páginas, mientras que la unión de Infoseek y Lycos produciría un índice de 90 millones de páginas y la de Excite y Northern Light alcanzaría los 230 millones, "otras evaluaciones realizadas con anterioridad han subestimado el tamaño de la web" ([Lawrence, 1998](#)). En el análisis del número de enlaces erróneos facilitados por cada motor, Hot Bot destaca por ser el que más proporciona (5%) frente a Lycos (1.6%), es decir, al menos en este estudio, el índice de tamaño más pequeño proporciona menos errores entre sus enlaces que el de mayor tamaño.

## **Gordon y Pathak (1999).**

Gordon y Pathak desarrollan un estudio donde analizan siete motores de búsqueda: Hot Bot, Magellan, Infoseek, Excite, Lycos, Alta Vista y Open Text; junto a un directorio, Yahoo ([Gordon, 1999](#)). Estos autores consideraban muy importante la naturaleza intrínseca de los motores, de manera que no veían viable extrapolar los tests desarrollados sobre los SRI tradicionales, al tener estos sistemas unas bases de datos mucho más estructuradas y una naturaleza informativa más homogénea que en el contexto de la web. Es por ello que abogan por la necesidad de emplear una serie de características más específicas, para complementar los parámetros tradicionales empleados en la evaluación de la recuperación de información.

En su experimento, Gordon y Pathak examinaron el comportamiento de la exhaustividad y precisión ofreciendo un importante número de conclusiones, entre las que se pueden destacar:

1. La efectividad media en la recuperación de información es baja, alrededor del 10% en los primeros diez documentos recuperados.
2. Se detectan diferencias estadísticas considerables entre la precisión de los motores analizados. Alta Vista, Open Text y Lycos resultan los mejores y Yahoo ocupa el último lugar. Otra cuestión importante a destacar es la amplia distancia entre el primero y el segundo (más de 20 puntos porcentuales).
3. En la exhaustividad no se detectan diferencias sustanciales.
4. La realización de una búsqueda aparece más fuertemente vinculada al algoritmo de localización de la información que a los tipos de búsqueda que permite.
5. Aunque los motores siguen, en líneas generales, los principios generales de la recuperación de información, su funcionamiento varía de alguna manera ya que es clara la tendencia a recuperar documentos irrelevantes o

lejanamente relevantes de algunos de ellos, hecho difícilmente producible en el caso de un sistema tradicional.

Gordon y Pathak aportan otras conclusiones no menos importantes: los índices de los motores poseen tamaños muy diferentes (algunos son diez veces más grandes que otros) y (aunque algún motor, en su publicidad afirme lo contrario), ninguno pretende realmente indexar toda la web. Asimismo, los motores también difieren en la actualización periódica de los datos, en la posibilidad de que los usuarios añadan páginas por su cuenta, en el plazo de tiempo para incorporar una nueva página indexada tras tener noticia de su existencia y en el seguimiento de la disponibilidad de los enlaces.

También estudian el solapamiento detectado bajo dos perspectivas, (a) el grado de solapamiento existente entre los documentos recuperados por los ocho motores analizados y (b) el grado de solapamiento existente entre los documentos recuperados por los ocho motores analizados, que han sido considerados relevantes. Resultan sorprendentemente bajos los resultados del primer tipo, inferior al 10% de solapamiento, manteniéndose constante cuando el análisis se prolonga a los 50, 100 y 200 primeros documentos recuperados. El segundo tipo de solapamiento presenta unos valores mucho mayores de coincidencia que el anterior, ya que el número de documentos relevantes recuperados por un único motor oscila entre el 22.45% y el 25.4%, es decir que el grado de solapamiento no desciende nunca del 70%. Esta serie de datos permite suponer que los motores muestran muchos documentos distintos unos de otros pero, en principio, un porcentaje importante de los relevantes forman parte de la respuesta de cada uno de ellos.

### **Gwizdka, J. and Chignell (1999).**

Estos autores ponen en duda la tradicional asignación binaria de los valores de la relevancia (documento relevante-documento no relevante), ya que un documento difícilmente será relevante o no relevante en términos absolutos. En su estudio analizan un considerable número de aspectos: la precisión (entendida en los términos de la función definida por Leighton y Srivastava), el alineamiento de los documentos (estableciendo una función diferencial de la precisión conforme aparezcan alineados), el esfuerzo del usuario, la longitud esperada de búsqueda, el número de enlaces erróneos y el número de enlaces duplicados.

Como resultados globales de este experimento, Alta Vista presenta mejores resultados que los otros dos motores analizados, en términos de precisión y de diferencial de precisión. El solapamiento es bajo, fundamentalmente porque los motores "emplean diferentes procedimientos de localización de los documentos y porque sus métodos de recopilación e indización son sustancialmente distintos" ([Gwizdka, 1999](#)). El experimento no detecta efectos significativos de la precisión entre los motores y los dominios geográficos analizados, siendo Alta Vista el que presenta mejores índices de cobertura en todos ellos.

### **Ming (2000).**

Este completo estudio analiza la precisión, el tiempo de respuesta del sistema, la interface de usuario, el número de aciertos e introduce un factor que denomina sensibilidad, factor sobre el cual incide la calidad de los enlaces devueltos por el motor de búsqueda. Por desgracia, este estudio sólo afecta a Yahoo, Alta Vista y Lycos, y ofrece como resultados más destacados el hecho de no encontrar diferencias significativas en el número de documentos devuelto por cada sistema; prefiere el interface y el tiempo de respuesta ofrecido por Yahoo (aunque reconoce que es una opinión subjetiva). En lo relacionado con el análisis de la precisión, destaca que el valor medio de los tres motores evaluados en los primeros diez documentos devueltos es ligeramente superior al obtenido cuando sólo se toman en cuenta los cinco primeros documentos devueltos. En relación al parámetro de la sensibilidad, Ming afirma que Yahoo supera a los otros dos, resaltando los pobres valores de Alta Vista ([Ming, 2000](#)).

### **Notess (2000).**

Dentro del campo de los estudios experimentales, este autor proporciona cálculos de solapamiento, acierto único y enlaces fallidos. El solapamiento detectado es ligeramente superior que en el estudio de Gordon y Pathak, debido al crecimiento de los índices y al alto número de motores analizados. El porcentaje de enlaces fallidos determinado se recoge en la Tabla 3, siendo ligeramente elevado en el caso del motor Alta Vista:

Motor	% enlaces fallidos	% error tipo 400
-------	--------------------	------------------

Alta Vista	13,7	9,3
Excite	8,7	5,7
Northern Light	5,7	2
Google	4,3	3,3
Hot Bot	2,3	2
All the Web (Fast)	2,3	1,8
MSN (Inktomi)	1,7	1
Answers	1,3	0,7

**Tabla 3:** Porcentajes de fallos en las referencias ofrecidas por los motores de búsqueda.

Fuente: Notess, G. R. Search Engine Statistics (En línea).  
Bozeman, MT: Notes.com, 2000.

<<http://www.searchengineshowdown.com/stats/dead.shtml>>

(Consulta: 12 Dic. 2002)

## Ljosland (1999-2000).

De los trabajos realizados en los últimos años, destacan especialmente los llevados a cabo por la investigadora noruega Ljosland. Resaltan principalmente dos contribuciones: la primera de ellas analiza las cuestiones más importantes a considerar en la evaluación de los motores de búsqueda, la segunda analiza el funcionamiento de un importante número de motores de búsqueda cuando se les interroga por palabras de uso poco corriente.

El primero de los trabajos, presentado en la Conferencia SIGIR'99 (acrónimo de la reunión "22th International Conference of Research and Development in Information Retrieval", celebrada en Berkeley), estudia la relevancia de los documentos recuperados, introduciéndose un nivel intermedio de relevancia: 0, 0.5 y 1 ([Ljosland, 1999](#)). Otro aspecto muy importante es el planteamiento previo que realiza la autora sobre cómo reaccionar ante situaciones muy frecuentes en la web, tales como encontrarnos en la respuesta de un motor dos páginas de igual contenido pero alojadas en sitios web distintos o mirrors; la autora se pregunta "¿son páginas relevantes ambas o sólo una?" ([Ljosland, 2000](#)). Los contenidos seguramente sí, pero la segunda página no va a ser de ningún interés para el usuario, así que su relevancia, para Ljosland es nula. Esta decisión, aunque susceptible de discusión y de posibles reinterpretaciones, no puede pasar desapercibida, en tanto que el análisis de la relevancia que se lleva a cabo en este estudio se encuentra estrechamente vinculado al contexto en el que se desarrolla: la web, considerando este no presente en la mayoría de estudios anteriores.

Es fundamentalmente, este aspecto del estudio de Ljosland, el que le confiere una mayor representatividad frente a otros estudios que simplemente trasladan los criterios empleados en los estudios realizados sobre los tradicionales SRI a la web si entrar a considerar situaciones de nuevo cuño.

La autora compara únicamente tres motores: Alta Vista, Google y All the Web, siendo el primero un motor de consolidada posición y los otros dos, nuevos proyectos que buscan hacerse un hueco dentro de este amplio conjunto. El experimento muestra que, cuando no se considera la relevancia parcial, se obtienen valores de precisión media de 0.4 para Alta Vista, 0.7 para Google y de 0.4 para All the Web. Si se considera la relevancia parcial, suben un poco todos los valores. Otro dato importante medido es la posibilidad de encontrar un documento relevante en el primer lugar de la lista de documentos devueltos por cada motor, siendo de nuevo a Google el motor que más destaca (80% frente al 50% del All the Web y el 30% obtenido por Alta Vista).

El segundo trabajo de Ljosland analiza la efectividad de la recuperación de información de veinte sistemas entre los que mezcla motores, directorios y meta buscadores. A todos les pide que localicen documentos con palabras de uso poco frecuente.

**Documentos  
recuperados**

**Documentos relevantes con  
duplicados**

**Documentos relevantes sin  
duplicados**

**Búsquedas no  
vacías**

Fast	93	74	70	9
Askjeeves	86	56	49	10
Northernlight	81	44	38	9
Inferencefind	69	53	41	9
Egosurf	66	50	41	7
Excite	66	47	35	9
Oingo	55	43	35	9
Alta Vista	50	45	38	9
MSN	47	37	27	8
Yahoo	40	34	28	8
Infotiger	39	36	29	9
Google	38	33	27	6
snap.com	38	30	22	8
Hot Bot	29	23	19	7
DirectHit	27	5	5	3
Infoseek	18	13	11	6
Lycos	12	11	10	3
Goto.com	9	6	6	2
Euroseek	6	6	6	4
Webcrawler	5	5	4	5

**Table 4:** Resultados del estudio de Ljosland de búsqueda de palabras raras. Fuente: Lojsslund, M. *Evaluation of twenty Web search engines on ten rare words ranking algorithms*. (En línea). Trondheim: Sør-Trøndelag University, 2000. <<http://www.aitel.hist.no/~mildrid/dring/paper/Comp20.doc>> (Consulta: 12 May. 2002)

La Tabla 4 recoge los documentos recuperados por cada motor y cuántos son relevantes (sin discernir documentos iguales que aparecen en sitios web distintos, que Ljosland denomina duplicados), la tercera columna elimina esos duplicados y en último lugar, se muestra el número de búsquedas que no han resultado vacías, es decir, que han devuelto al menos un documento. Lojsslund no emplea aquí la "relevancia parcial" (Ljosland, 2000).

Los resultados muestran que tanto Fast como AskJeeves devuelven una mayor cantidad de documentos y poseen mejor precisión. El estudio de la precisión permite deducir que es mejor para los motores que poseen escasa cobertura y que pocas veces presentan listas vacías de documentos (es decir, que no devuelven nada). La Tabla 5 resume los resultados obtenidos en ambos ensayos:

Cobertura	Precisión	Exhaustividad	Exhaustividad única
Askjeeves (0,39)	Euroseek, Webcrawler (1,00)	Askjeeves (0,43)	Fast (0,58)
Fast (0,34)		Fast (0,42)	Askjeeves (0,57)
InferenceFind (0,28)	Lycos (0,95)	InferenceFind (0,38)	InferenceFind (0,46)

**Table 5:** Tabla resumen del estudio de Ljosland de búsqueda de palabras raras ordenados por la precisión. Fuente: Lojsslund, M. *Evaluation of twenty Web search engines on ten rare words ranking algorithms*. (En línea). Trondheim: Sør-Trøndelag University, 2000. <<http://www.aitel.hist.no/~mildrid/dring/paper/Comp20.doc>> (Consulta: 12 May. 2002)

Donde Ljosland resulta verdaderamente innovadora es en comparar los resultados devueltos por los distintos motores por medio de una función de similitud, la de Jaccard más concretamente ([SAL, 1983](#)). Así pretende averiguar hasta qué punto los conjuntos de documentos devueltos por algunos motores son realmente subconjuntos de los devueltos por otros. Así, más allá de las cifras obtenidas, en este segundo estudio destaca la originalidad del cálculo de la similitud de los resultados devueltos por cada motor, ya que la coincidencia en la composición de los índices de documentos de los distintos motores se ha venido calculando normalmente con base en el porcentaje de solapamiento y de referencia única, siendo esta nueva medida algo más elaborada, ya que además de determinar el grado de intersección de la respuesta de dos motores mide la similitud de la misma en cuando a la posición que ocupan los documentos en ella.

Con esta última aportación, no hay duda alguna sobre que ambos trabajos de Ljosland pueden considerarse de los más completos de los realizados hasta ahora en este campo, aún cuando los resultados no resultan fácilmente extrapolables por dos razones: en el conjunto de SRI analizados se entremezclan motores, directorios y meta buscadores, y en segundo lugar, las cuestiones planteadas se refieren a la presencia de términos de uso poco frecuente en documentos, no a la relevancia o no relevancia de un documento con un tema objeto de una necesidad de información.

## C|NET search site olympics (Thomas, 2002).

Thomas, publica en la web de la revista electrónica C|NET el resultado de un evaluación de cinco motores de búsqueda (Alta Vista, Excite, Google, Lycos y MSN Search), realizada a principios del año 2002. En la misma medía dos conceptos, uno experimental y otro más subjetivo (el diseño artístico del sitio web y la presentación de los resultados). El criterio experimental era la relevancia de los enlaces devueltos. Los resultados otorgan mejores resultados a Google con una amplia ventaja sobre el resto.

## Sinopsis de estudios experimentales.

La revisión de los estudios de evaluación de los motores de búsqueda de tipo experimental ha recopilado un total de catorce trabajos, que han empleado, a lo largo de su desarrollo, diecisiete variables diferentes para llevar a cabo sus procesos de análisis. Estos estudios se han sintetizado en la Tabla 6.

	PB	TR	GI	DO	TI	P	E	CO	FA	PO	HU	EE	ED	FO	SO	LO	EN	
Chu y Rosenthal	X	X	X	X		X	X	X	X	X								Alta Vista
Leighton, 1995						X	X				X	X	X					Lycos
Leighton, 1999						X												Alta Vista
Arents			X		X	X								X				Google
Bharat					X										X			Alta Vista
Ralph					X	X	X	X						X				Hot Bot
Wishard						X												Lycos
Lawrence					X			X				X			X			Hot Bot
Gordon					X	X	X		X						X			Alta Vista
Gwidzka			X			X						X	X	X		X		Alta Vista

Ming	X	X	X	X					X	Alta Vista
Notess			X		X		X	X		Fast
Lojlsland, 1999				X						Google
Lojlsland, 2000			X	X	X		X		X	Fast
Thomas, 2002				X					X	Google

**Tabla 6:** Sinopsis de los estudios experimentales analizados. PB: posibilidades de búsqueda. TR: tiempo de respuesta. GI: Interface gráfica de usuario. DOC: documentación. TI: tamaño del índice. P: precisión. E: exhaustividad. CO: cobertura. FA: frecuencia de actualización. PO: porción de página indexada. FID: fidelidad. HU: acierto único. EE: enlaces erróneos. ED: enlaces duplicados. FO: formato de presentación y alineamiento. SO: solapamiento LO: longitud esperada de búsqueda AL: alineamiento EN: calidad de los enlaces

El número de estudios realizados es menor que en el caso de los explícitos y se han usado más variables, ocurrencias perfectamente comprensibles ya que un estudio experimental suele ser más complejo en su elaboración. El número medio de variables empleadas por trabajo alcanza el valor de 4.21, resultado de dividir las 59 variables entre los 14 estudios, duplicando así la cifra obtenida en los estudios explícitos. Al igual que en la presentación de los estudios explícitos, en la última columna de la tabla se indica qué motor recibe la mejor valoración en cada estudio. En esta ocasión aparecen citados cinco motores, dos menos que en el caso anterior. Las características más veces empleadas son: precisión (12 veces), tamaño del índice (8 veces), enlaces erróneos y solapamiento (5 veces). La mayoría de los estudios que emplean la precisión, muestran a Alta Vista como el mejor. Con el tamaño del índice no existe una tendencia clara, ya que se proponen hasta cuatro motores distintos. La presencia de Google en los estudios (en los más recientes o actualizados) suele implicar su designación como el de índice más grande, en cambio, cuando Google no aparece (seguramente por la fecha del estudio, anterior a su puesta en marcha), Alta Vista ocupaba esa posición de privilegio.

El número de motores analizados en cada estudio y el número de variables empleadas es otro parámetro de interés, ya que sirve para valorar la relevancia de estos trabajos. En la Tabla 7 se observa cómo abordan los autores sus análisis:

	Número de motores analizados						Número de variables utilizadas
	<=3	4-6	7-9	10-15	15-20	>20	
Chu y Rosenthal	X						9
Leighton, 1995		X					5
Leighton, 1999		X					1
Arents				X			4
Bharat		X					2
Ralph			X				5
Wishard					X		1
Lawrence		X					4
Gordon			X				5
Gwidzka	X						6
Ming	X						5



Notess		X	5
Lojssland, 1999	X		1
Lojssland, 2000		X	5
Thomas, 2002	X		2

**Tabla 7:**Síntesis del número de características empleadas en los estudios experimentales.

La mayoría de las veces se analizan, como máximo, seis motores (9 de 14 veces). En cuanto al número de parámetros de cada estudio, el de Chu y Rosenthal (curiosamente el primero) llega a emplear nueve, siendo más normal analizar cinco o seis variables (hecho ocurrido 7 de las 12 ocasiones restantes). Siguiendo la línea marcada por este estudio, al analizar un número considerable de parámetros en pocos motores, lo normal sería que este grupo mayoritario de estudios analizase un conjunto reducido de motores, aunque no existe una tendencia clara en este punto, ya que su cobertura es muy diversa.

## Propuestas globales de evaluación.

Oppenheim presenta una sugerencia de criterios mínimos necesarios para el diseño de una metodología de evaluación, fruto de una exhaustiva síntesis de las medidas empleadas en otros trabajos anteriores (Oppenheim, 2000). A este trabajo le sigue un interesante estudio realizado por Savoy y Picard ([Savoy, 2001](#)), donde analizan la efectividad de los distintos modelos sobre los que se basan los SRI en la web, en lugar de analizar el comportamiento de un motor específico frente a otro. El tercero de los trabajos que se presenta expone la necesidad de encontrar una metodología ajena a los juicios de relevancia, basada en unos parámetros de sensibilidad y utilidad de los documentos. Por último, el cuarto trabajo recogido presenta una propuesta global de evaluación de los SRI en la web elaborada desde el punto de vista del usuario final ([Johnson, 2001](#)).

### Oppenheim (2000).

A partir de la síntesis de estudios de evaluación anteriores, los autores formulan una metodología de evaluación de los motores de búsqueda, agrupando los métodos más empleados en cuatro categorías:

- Evaluaciones a pequeña escala
- Evaluaciones basadas en los tests Cranfield
- Evaluaciones basadas en los tests Cranfield con estimación del tamaño del motor
- Evaluaciones que eluden la exhaustividad

La Tabla 8 sintetiza el conjunto de criterios empleados y las conclusiones que de su estudio se extrajeron.

Criterio	Conclusiones generales
Número de páginas cubiertas y cobertura	Ningún motor de búsqueda por sí solo indexa toda la web. Los resultados además proyectan algunas dudas sobre la validez de usar la exhaustividad relativa como medida.
Actualización del índice y número de enlaces erróneos	La frecuencia de actualización es un parámetro casi tan importante como el tamaño del índice del motor.
Relevancia	Normalmente se hace uso de una visión binaria de la relevancia (sí-no), aunque otros introducen escalas.
Sintaxis	Se identifican tres categorías: frases, lenguaje natural y booleanas.
Materias	Campo no muy estudiado. Es normal que un motor ofrezca mejores resultados en un área que en otra.

La dinámica naturaleza de la web	Una página puede cambiar o desaparecer al poco de ser indexada. Esto implica que los resultados ofrecidos por distintos motores sean diferentes. Los experimentos deben contemplar este dinamismo y las evaluaciones deben hacerse en fechas próximas y repetirse a para incorporar los cambios que se vayan produciendo.
Tiempo de respuesta	Parámetro de difícil cálculo y muy supeditado al tráfico de la red
Características diferentes del sistema	Las características de cada sistema influyen considerablemente en sus usuarios. Destacan varios trabajos que recomiendan hacer uso de motores <i>ad hoc</i> para cada necesidad particular de información.
Opciones de búsqueda	Si los motores hacen uso de búsquedas simples y/o avanzadas. Algunos emplean otros tipos de búsquedas más complejas. También se estudian las preferencias de búsqueda de los usuarios
Factores humanos y cuestiones de la interface	Los motores deben evaluarse bajo una perspectiva de usuario que evite problemas con la subjetividad. El análisis de la interface lleva implícito una obsolescencia. Estos ingenios resultan inaccesibles, en gran medida, para invidentes.
Calidad de los resúmenes	Aspecto que parecía relegado a un papel residual dentro de la evaluación de los motores. Se han comparado resúmenes procedentes de motores y directorios con los de bases de datos en línea de amplia difusión, comparando la longitud y legibilidad. El resultado indica que los directorios poseen mejores resúmenes

**Tabla 8:** Resumen de características empleadas en las evaluaciones de los SRI. Fuente: [Oppenheim, 2000](#).

## Savoy y Picard (2001).

Este estudio investiga "si las técnicas usadas en los SRI mejoran la efectividad en la recuperación de información cuando se aplican a una colección de documentos web " ([Savoy, 2001](#)), poniendo en duda la aplicación de técnicas tan comunes como el aislamiento de la base de las palabras , la presencia de una lista de palabras vacías o la presencia de palabras significativas en el título de la página web.

Savoy y Picard miden la precisión tras analizar los 10 primeros documentos devueltos por cada modelo y los 20 primeros documentos. Este interesante experimento se desarrolló sobre la base de datos WT2g, colección mantenida por TREC con un tamaño de 2194 MB y que consta de 247.491 páginas distintas extraídas de 969 URLs, con más de un 1.850.000 términos de indización. Los resultados de este primer análisis son los siguientes:

- El modelo probabilístico devuelve unos resultados ligeramente superiores a los obtenidos por el modelo del vector clásico.
- El modelo vectorial que emplea el esquema tf-idf no arroja resultados muy satisfactorios.
- Resultan especialmente pobres los resultados del esquema basado en la frecuencia de aparición de los términos en los documentos.
- La precisión medida desde el punto de vista de promedio resulta muy baja tras analizar una serie de mil documentos (en torno al 25% en los casos más precisos).

Un dato a considerar es el considerable descenso de los valores de la precisión cuando se hace uso de los términos recogidos en los títulos de las páginas web y en las zonas de descripción de los documentos. Por lo tanto, en contra de una de los consejos más clásicos que se realizan a la hora de editar una página web, el insertar una breve descripción de la página en el título, no mejora significativamente la precisión en la recuperación de información. Igualmente, asignar un mayor peso a las palabras del título o de los encabezados de las páginas, no posee efectos significativos sobre la precisión.

Casi de forma paralela a la realización de este estudio, estos mismos autores analizaron la efectividad de tres modelos de recuperación de información en la web: el clásico (representado por el TRECEval software y que siguen la mayoría de los motores), el de extensión de enlaces (el usado por Google) y uno basado en el modelo probabilístico ([Picard, 2001](#)). Aunque tampoco comparan motores de búsqueda en particular, sí evalúan las prestaciones de Google frente a los motores clásicos. Los valores medios obtenidos, cuando se han considerado los enlaces que salen de las páginas analizadas, se recogen en la Tabla 9:

TRECeval	Extensión enlaces	Probabilístico	Probabilístico vs vs TRECeval	Probabilístico vs Extensión Enlaces
0,253	0,266	0,267	+5.53%	+0.37

**Tabla 9:** Resumen de resultados de la efectividad de los distintos modelos de recuperación de información en la web. Fuente: *Picard, J, and Savoy, J. Searching and Classifying the web using hyperlinks: A logical approach.* (En línea) Neuchatel: III, Université, 2001. <<http://www.unine.ch/info/Gi/Papers/ictai.pdf>> (Consulta: 21 Nov. 2002)

## Schlichting y Nilsen (1997).

Para los autores de este trabajo, las primeras evaluaciones que se realizaron de los motores de búsqueda "poseían un escasísimo nivel científico". De hecho, Schlichting y Nilsen indican que los trabajos de Winsip y de Leighton de 1995 constituyen un avance importantísimo en este campo de trabajo. El objetivo de este trabajo es mostrar una metodología que pueda emplearse para comparar motores de búsqueda y otro tipo de posibles futuros sistemas inteligentes de recuperación de información, metodología que debe encontrar alguna manera de medir la calidad de los resultados ofrecidos por un motor. Los autores proponen la Metodología SDA, que proporciona dos medidas:  $d'$  que mide la sensibilidad del motor de búsqueda en hallar información útil y  $\beta$  que mide cómo de conservador (o de liberal) es el comportamiento de motor a la hora de determinar qué páginas deben formar parte de la respuesta (mide el grado de flexibilidad del motor a la hora de considerar relevante un nuevo documento analizado). Este método "supera el alcance del simple cómputo de los aciertos, que sólo emplea una pequeña cantidad de los datos e ignora el amplio contexto de la búsqueda de información" ([Schlichting, 2001](#)).

Los motores analizados fueron Alta Vista, Lycos, Infoseek y Excite. Se obtuvieron un total de 200 documentos, de los cuales sólo 54 fueron considerados relevantes, siendo Lycos el que obtuvo un mayor número. El primer paso de esta metodología es asignar una categoría a cada uno de los resultados devueltos con base en la relevancia de cada uno (acierto, falsa alarma, perdido y rechazado correctamente). En la Tabla 10 aparecen los resultados de estas ratios obtenidas en este experimento:

TRECeval	Extensión enlaces	Probabilístico	Probabilístico vs vs TRECeval	Probabilístico vs Extensión Enlaces
0,253	0,266	0,267	+5.53%	+0.37

**Tabla 10:** Ratios de acierto y falsa alarma en cuatro motores de búsqueda. Fuente: *Signal Detection Analysis of WWW Search Engines.* (En línea) Seattle: Microsoft, 1997. <<http://www.microsoft.com/usability/webconf/schlichting/schlichting.htm>> (Consulta: 21 Nov. 2002)

Esta metodología no se detiene en este punto, ya que estos valores le sirven para la determinación de la medida de la sensibilidad ( $d'$ ) y de la flexibilidad ( $\beta$ ). Los resultados obtenidos en el estudio asignaron a Lycos (con un discreto valor de 0.48 sobre 2, aunque muy por encima del resto), como el motor más sensible. El más claro resultado de este estudio es que el rendimiento de los motores de búsqueda "está muy lejos de ser considerado ideal" ([Schlichting, 2001](#)).

## Johnson, Griffiths y Hartley (2001).

En la *Universidad Metropolitana de Manchester*, se ha desarrollado recientemente un estudio orientado a fijar un marco global para la evaluación de los motores de búsqueda. Tras una descripción de estos sistemas y una presentación de su evolución, el estudio revisa las evaluaciones previas desde una perspectiva sistémica que evoluciona hacia un punto de vista centrado en el usuario. En último lugar, se construye un marco global para la evaluación basado en criterios de satisfacción del usuario, las medidas que se van a emplear y el contexto donde se desarrollará la evaluación. Esta propuesta se esquematizarse en la Tabla 11:

**Evaluación de la satisfacción del usuario de los SRI en la web.  
(Johnson, Griffiths y Hartley, 2001)**

<b>Eficacia</b>	<b>Eficiencia</b>	<b>Utilidad</b>	<b>Interacción</b>
P1: <i>precisión</i>	Tiempo empleado en la sesión de búsqueda	Valor global de los resultados	Satisfacción del usuario con el formato de salida de los resultados
P2: satisfacción del usuario con la <i>precisión</i>	Tiempo de respuesta del sistema	Satisfacción con los resultados	Satisfacción del usuario con las posibilidades de manejo de los resultados
P3: comparación de P1 y P2	Tiempo de establecimiento de la <i>relevancia</i>	Grado de resolución del problema	Satisfacción del usuario con la visualización de la representación de los documentos
R1: el alineamiento de los documentos que hace el sistema		Valor medio de la participación	I1: Satisfacción del usuario con la interface del sistema
R2: satisfacción del usuario con el alineamiento R1		Calidad de las fuentes	I2: Satisfacción del usuario con el formato de la pregunta
R3: comparación de R1 y R2		Números de enlaces seguidos	I3: Satisfacción del usuario con las posibilidades de modificar el formato de la pregunta  I4: Satisfacción del usuario con la claridad de la vista global del sistema en la realización de una pregunta

**Tabla 10:** Propuesta de un marco global de evaluación de motores de búsqueda del *Informe DEVISE*.

Las medidas de la efectividad del sistema se obtienen a partir de una serie de entrevistas realizadas a los usuarios del sistema, una vez realizadas las búsquedas. Los usuarios valoran tanto la precisión como el alineamiento de los documentos. La eficiencia del sistema se mide en términos de tiempo, aunque en este caso se diferencia entre el tiempo de respuesta del sistema (el parámetro más utilizado de este grupo en otros estudios anteriores), el tiempo que necesita el usuario para la recuperación de información y el tiempo que precisa el usuario para discernir la validez del documento recuperado.

La utilidad de los documentos se mide en términos de valor medio de la participación de los usuarios, a partir de sus propias opiniones sobre el grado en el que la información recuperada contribuyó a satisfacer sus necesidades de información. La interacción con el sistema, "este concepto es a menudo citado pero pocas veces ha sido definido. En el contexto de los motores de búsqueda hace referencia a cómo el usuario directamente interacciona con el sistema para recuperar la información que necesita" ([Johnson, 2001](#)).

En la mayoría de los estudios previos, la interactividad se establece por medio de la medida I1 (satisfacción del usuario con la interfase del sistema), pero Johnson, Griffiths y Hartley destacan la importancia de esta interactividad en el momento de la realización de una pregunta, para ello definen I2 (satisfacción del usuario con el formato de la pregunta), I3 (satisfacción del usuario con las posibilidades que el sistema le ofrece para modificar la pregunta sin tener que repetir toda la consulta) y por último, la medida I4 (satisfacción del usuario con la claridad de la vista global del sistema en el momento de realización de una pregunta). También se establecen diferencias entre el formato de visualización de los documentos y el formato de manejo de los documentos, conceptos que tradicionalmente siempre han sido considerados como uno sólo.

Con lo cual, este estudio, viene a representar de forma empírica, las principales necesidades de un usuario de un SRI, el usuario necesita información de valor que cubra sus necesidades de información y la misma le debe ser entregada en un espacio de tiempo pequeño.

## Crítica.

Uno de los principales problemas de las evaluaciones (de cualquier tipo) efectuadas sobre los SRI en la web reside en el hecho de que la mayoría se centran en analizar la efectividad del acceso físico a los datos, cuando lo que verdaderamente cobra importancia hoy en día, es analizar el comportamiento del acceso lógico a los datos, esto es, el contenido informativo de la recuperación. Si bien se ha desarrollado un amplio conjunto de medidas destinadas a llevar a cabo estas evaluaciones, la mayor parte de los estudios continúan empleando los juicios basados en la relevancia de los documentos para valorar esta efectividad. Destacan sobremanera la precisión y la exhaustividad (pares de valores E-P). La primera de estas medidas no presenta problemas para su determinación, pero la exhaustividad es una medida de difícil e indeterminado cálculo. Es por ello que los cálculos de la efectividad basados en estos pares de valores, no permiten una aproximación exacta, sino aproximada o relativa. A este problema debe añadirse que los pares de valores E-P no ofrecen una visión global por si solos, habiéndose definido un amplio conjunto de medidas de valor simple (basadas, generalmente, en estos pares de valores), para medir la efectividad.

La necesidad intrínseca de evaluación de los SRI se ha trasladado al nuevo contexto de la web, surgiendo trabajos coetáneos al desarrollo de los primeros directorios y motores de búsqueda. Al mismo tiempo se han trasladado las dudas subyacentes sobre los métodos empleados en estas evaluaciones. Otra situación que contribuye a aumentar estas dosis de incertidumbre es la especial naturaleza de la web, que propicia situaciones que en el entorno de los SRI tradicionales resultarían inverosímiles, propiciando que estos sistemas no estén siendo correctamente evaluados, ya que no pueden emplearse exclusivamente las medidas basadas en la relevancia, sino que resultan necesarias medidas específicas.

Existen profundas divergencias entre los métodos empleados en las evaluaciones de motores y directorios. Los trabajos que analizan los aspectos explícitos no sirven para representar, en modo alguno, la efectividad de la recuperación de información. Los estudios experimentales actualizan los criterios empleados en los SRI clásicos al entorno de la web y representan mejor la naturaleza dinámica de este nuevo contexto y el funcionamiento de los sistemas analizados, aunque realmente las aportaciones a este campo son escasas.

Entre este restringido conjunto de contribuciones destaca el diseño de una función de precisión que considere la posición de los documentos en la respuesta, propuesta por Leighton y Srivastava, y el empleo de una función de similitud (Ljosland propone la función de Jaccard) para determinar en qué porcentaje las respuestas de un motor forman un subconjunto de las respuestas de otro. Con este cálculo se establece un nuevo método de determinación del solapamiento entre los índices de los motores de búsqueda, tomando como base las respuestas que estos motores llevan a cabo a las mismas preguntas, sondeando la analogía existente entre los conjuntos de documentos gestionados por estos sistemas y el orden en el que los mismos se presentan al usuario.

Los trabajos centrados en el establecimiento de un marco global de evaluación de los motores de búsqueda son sustancialmente distintos y resulta complicado hallar una serie de conclusiones comunes entre ellos. De su análisis se desprende el convencimiento de la necesidad de aplicar un enfoque multidimensional a la evaluación de la efectividad de la recuperación de información en la web, donde exista un cierto grado de relación entre las dimensiones analizadas. Estas metodologías propuestas resultan de difícil aplicación por dos razones: la complejidad del cálculo de las medidas propuestas y, en segundo lugar, el considerable nivel de abstracción de alguna de estas medidas, que las alejan del usuario convencional de la web.

## Conclusiones.

Para obtener esta adecuada visión, han de evaluarse conjuntamente motores de búsqueda de características similares, tanto en el tamaño de sus índices como en sus posibilidades de búsqueda. Del mismo modo, se deben tener en cuenta las restricciones geográficas que, en algunos casos presentan los motores, siendo más conveniente cotejar motores sin ningún tipo de restricción. La selección del idioma también es importante, en tanto que algunos motores prácticamente no pueden ser susceptibles de evaluación en un idioma distinto del inglés. Resulta esencial, antes de proceder a cualquier proceso de evaluación, fijar debidamente la serie de procedimientos a seguir cuando se presenten casuísticas típicas de la web, tales como recuperación de páginas de contenido similar alojadas en distintos sitios, muchas referencias al mismo sitio web (situación muy frecuente en el caso de los portales dinámicos), redireccionamientos, etc.



Otra circunstancia que debe clarificarse de antemano es la escala de determinación de la relevancia, contemplándose la posibilidad de emplear criterios de relevancia parcial. Creemos que sí es posible hablar de un comportamiento típico de los motores de búsqueda, independientemente de los mejores o peores resultados que ofrezcan en los diversos experimentos realizados, ya que los mismos, seguramente, se deben más al espacio sobre el cual se definen las vistas lógicas de los documentos y las necesidades de información (de naturaleza difusa y escasamente similar) que a las diferencias sustanciales que pueden existir entre un sistema y otro. Finalmente, consideramos que las próximas propuestas de evaluación de los motores de búsqueda deben abocarnos a encontrar una forma de evaluación que elimine la toma de decisiones, que casi siempre añaden un elemento de subjetividad al experimento, y por otro lado permita que de manera simple puedan analizarse muestra de resultados suficientemente significativas (en cuanto a su volumen) y en periodos cortos en el tiempo. Es en esa dirección donde en la actualidad venimos investigando posibles soluciones.

## Referencias.

- Arents, H. (1997) [A rated selection of Internet search tools](http://www.lib.rug.ac.be/internet/search.html) Gent: University. <http://www.lib.rug.ac.be/internet/search.html> (12 feb. 2002)
- Bharat, K. and Broder, A. (1998) "[A technique for measuring the relative size and overlap of public web search engines](#)". Paper delivered to the Seventh international World Wide Web conference, Brisbane, Australia, 14-18 April, 1998. <http://www7.scu.edu.au/programme/fullpapers/1937/com1937.htm> (21 dic. 2002)
- Bharat, K. and Broder, A. (1998) [Measuring the Web](#). Palo Alto, CA: Digital Systems Research Center. <http://www.research.compaq.com/src/whatsnew/sem.html> (21 Dic. 2002)
- Chang, G., Healey, M.J., McHugh, J.A.M. y Wang, J.T.L. (2001) *Mining the World Wide Web: an information search approach*. Norwell, MA: Kluwer Academic Publishers, 2001.
- Chu, H. and Rosenthal, M. (1996) [Search engines for the World Wide Web: a comparative study and evaluation methodology](#). Paper presented at the ASIS 1996 Annual Conference, October 19-24, 1996. <http://www.asis.org/annual-96/ElectronicProceedings/chu.html> (4 may. 2002)
- Davis, E.T. (1996) [A comparison of seven search engines](#). Columbus, OH: Kent State University. <http://www.iwaynet.net/~lsci/Search/paper.htm> (4 abr. 2002)
- Delgado Domínguez, A. (1998) [Mecanismos de recuperación de Información en la WWW](#) Mallorca: Universitat Illes Balears, 1998. (16 Jun. 2002) <http://dmi.uib.es/people/adelaide/tice/modul6/memfin.pdf> (15 may. 2002)
- Gordon, M. & Pathak, P. (1999) "Finding information on the World Wide Web: the retrieval effectiveness of search engines." *Information Processing and Management* **35**, 141-180
- Gwizdka, J. & Chignell, M. (1999) [Towards information retrieval measures for evaluation of Web search engines](#). Toronto: University of Toronto, Mechanical & Industrial Engineering Department. (Unpublished manuscript) [http://www.imedia.mie.utoronto.ca/people/jacek/pubs/webIR\\_eval1\\_99.pdf](http://www.imedia.mie.utoronto.ca/people/jacek/pubs/webIR_eval1_99.pdf) (21 mar. 2002)
- Johnson, F.C., Griffiths, J.R. and Hartley, R.J. (2001) [DEVISE: a framework for the evaluation of Internet search engines](#). London: Re:source - The Council for Museums, Archives and Libraries. <http://www.mmu.ac.uk/h-ss/cerlim/projects/devise/devise-report.pdf> (21 dic. 2002)
- Lancaster, F. W. & Warner, A.J. (1993) *Information retrieval today*. Arlington, VA: Information Resources.
- Lawrence, S. & Giles, C.L. (1998) "[Searching the World Wide Web](#)". *Science*, **280**,(5360), 98-100. <http://www.neci.nec.com/~lawrence/science98.html> (21 dic. 2002)
- Lebedev, A. (1997) [Best search engines for finding scientific information in the Web](#). Moscow: State University. <http://www.pharm.unito.it/itcrs/comparis.html> (21 dic. 2001)
- Leighton, H. V. (1995) [Performance of four World Wide Web \(WWW\) index services: Infoseek, Lycos, WebCrawler, and WWWorm](#). Winona, MN: Winona State University. <http://www.winona.msus.edu/library/webind.htm> (15 dic. 2002)
- Leighton, H.V. & Srivastava, J. (1997) [Precision among World Wide Web search services \(search engines\): Alta Vista, Excite, Hotbot, Infoseek, Lycos](#). Winona, MN: Winona State University. <http://www.winona.msus.edu/library/webind2/webind2.htm> (15 dic. 2002)
- Leighton, H.V. and Srivastava, J. (1999) "First 20 precision among World Wide Web search services (search engines)." *Journal of the American Society for Information Science* **50**(10), 870-881.
- Ljosland, M. (2000) [Evaluation of Web search engines and the search for better ranking algorithms](#). Paper presented at the SIGIR99 Workshop on Evaluation of Web Retrieval August 19, 1999. <http://www.aitel.hist.no/~mildrid/drng/paper/SIGIR.html> (4 Dic. 2002)
- Ljosland, M. (2000) [Evaluation of twenty Web search engines on ten rare words ranking algorithms](#).



- <http://www.aitel.hist.no/~mildrid/drting/paper/Comp20.doc> (18 dic. 2002)
- Martínez Méndez, F.J. (2001) "[Aproximación general a la evaluación de la recuperación de información por medio de los motores de búsqueda en Internet](#)". *Scire*, 6(1), <http://www.um.es/gtiweb/fjmm/ibersid2000.PDF> (6 may. 2002)
  - Ming, H. (2000) [Comparison of three search engines](#). Toronto: University of Toronto. [http://gypsy.rose.utoronto.ca/people/ming/Three\\_Internet.pdf](http://gypsy.rose.utoronto.ca/people/ming/Three_Internet.pdf) (12 dic. 2002)
  - NetRatings//Nielsen. (2001) [Top 25 parent companies](#). Stanford, CT: NetRatings Inc. & Nielsen Media Research. <http://pm.netratings.com/nnpm/owa/NRpublicreports.toppropertiesmonthly> (13 nov. 2002)
  - Notess, G.R. (1999) [Comparing Internet search engines](#). In: *Strategies for the next millennium: proceedings of the Ninth Australasian Information Online & On Disc Conference and Exhibition*, Sydney, Australia 19-21 January 1999. <http://www.csu.edu.au/special/online99/proceedings99/103a.htm> (12 may. 2001)
  - Notess, G.R. (2002a) [Search engine statistics](#). Bozeman, MT: Notess.com. <http://www.searchengineshowdown.com/stats/> (4 dic. 2002)
  - Notess, G.R. (2002b) [Search engine database change over time](#). Notess.com. <http://www.searchengineshowdown.com/stats/change.shtml> (4 dic. 2002)
  - Notess, G.R. (2002c) [Search engine statistics: dead reports](#). Notess.com. <http://www.searchengineshowdown.com/stats/dead.shtml> (4 dic. 2002)
  - Notess, G.R. (2002d) [Search engine statistics: database total size estimates](#). Notess.com. <http://www.searchengineshowdown.com/stats/sizeest.shtml/v> (12 Dic. 2002)
  - Oppenheim, C., Morris, A., McKnight, C. & Lowley, S. (2000) "The evaluation of WWW search engines". *Journal of Documentation*, 56(2), 190-211
  - Peterson, R.E. (1997) ["Eight internet search engines compared."](#) *First Monday*, 2(2) [http://www.firstmonday.dk/issues/issue2\\_2/peterson/](http://www.firstmonday.dk/issues/issue2_2/peterson/) (12 Mar. 2002)
  - Picard, J. & Savoy, J. (2001) [Searching and classifying the web using hyperlinks: a logical approach](#). Paper delivered to the 23rd European Colloquium on Information Retrieval Research, 2001. <http://www.unine.ch/info/Gi/Papers/ictai.pdf> (18 Dic. 2002)
  - Salton, G. & McGill, M.J. (1983) *Introduction to modern information retrieval*. New York, NY: Mc Graw-Hill.
  - Savoy, J. & Picard, J. (1991) "Retrieval effectiveness on the Web". *Information Processing and Management*, 37(4), 543-569.
  - Schlichting, C. & Nilsen, E. (1997) [Signal detection analysis of WWW search engines](#). Seattle: Microsoft. <http://www.microsoft.com/usability/webconf/schlichting/schlichting.htm> (11 Dic. 2002)
  - SearchEngineWatch.com (2001) [Other global search engines](#) Jupitermedia Corporation. <http://www.searchenginewatch.com/links/otherglobal.html> (11 Dic. 2002)
  - SearchEngineWatch.com (2002) [The major search engines](#) Jupitermedia Corporation. <http://www.searchenginewatch.com/links/major.html> (11 Dic. 2002)
  - Slot, M. (1996) [Web matrix: overview matrix](#). Matt Slot. <http://people.ambrosiasw.com/~fprelect/matrix/overview.html> (1 Dic. 2001)
  - Stobart, S. & Kerridge, S. (1996) [WWW search engine study](#). Sunderland: University of Sunderland. <http://osiris.sunderland.ac.uk/sst/se/> (1 Dic. 2001) [Page not found 21 Dic. 2002 - see, for an alternative: "[An investigation into World Wide Web search engine use from within the UK - preliminary findings](#)", by S. Stobart and S. Kerridge. *Ariadne*, No. 6, 1996. <http://www.ariadne.ac.uk/issue6/survey/> (22 Dic. 2002)]
  - Sullivan, D. (2002) [StatMarket search engine ratings](#). Jupitermedia Corporation. <http://www.searchenginewatch.com/reports/statmarket.html/> (11 Dic. 2002)
  - Sullivan, D. (2001a) [Search engine sizes](#). Jupitermedia Corporation. <http://www.searchenginewatch.com/reports/sizes.html> (11 Dic. 2002)
  - Sullivan, D. (2001b) [Jupiter Media Metrix search engine ratings](#). Jupitermedia Corporation. <http://searchenginewatch.com/reports/mediamatrix.html> (11 Dic. 2002)
  - Sullivan, D. (2001c) [Nielsen//NetRatings search engine ratings](#) <http://searchenginewatch.com/reports/netratings.html> (11 Dic. 2002)
  - Sullivan, D. (2001d) [Searches per day](#). Jupitermedia Corporation. <http://www.searchenginewatch.com/reports/perday.html> (11 Dic. 2002)
  - Thomas, S. (2002) [5 popular search tools battle for the medal](#). CNET.com. <http://www.cnet.com/software/0-352105-8-8741537-1.html?tag=st.sw.352105-8-8741537-9.back2352105-8-8741537-1> (16 May. 2002)
  - Westera, G. (2001) [Comparison of search engine user interface capabilities](#). Gillian Westera personal page. <http://lisweb.curtin.edu.au/staff/gwpersonal/compare.html> (12 Abr. 2002)
  - Winship, I. ["World Wide Web searching tools - an evaluation"](#). *Vine* No. 99, 49-54.

<http://gti1.edu.um.es:8080/javima/World-Wide-Web-searching-tools-an-evaluation.htm> (4 May. 2002)

- Wishard, L. (1998) "[Precision among Internet search engines: an earth sciences case study](#)." *Issues in Science and Technology Librarianship*, No. 18. <http://www.library.ucsb.edu/istl/98-spring/article5.html> (4 May. 2001)
- Zorn, P., Emanoil, M., Marshall, L. & Panek, M. (1996) "[Advanced searching: tricks of the trade](#)." Online, 20(3) <http://www.infotoday.com/online/MayOL/zorn5.html> (19 Dic. 2002)

---

### Abstract

*A considerable number of proposals for measuring the effectiveness of information retrieval systems have been made since the early days of such systems. The consolidation of the World Wide Web as the paradigmatic method for developing the Information Society, and the continuous multiplication of the number of documents published in this environment, has led to the implementation of the most advanced, and extensive information retrieval systems, in the shape of web search engines. Nevertheless, there is an underlying concern about the effectiveness of these systems, especially when they usually present, in response to a question, many documents with little relevance to the users' information needs. The evaluation of these systems has been, up to now, dispersed and various. The scattering is due to the lack of uniformity in the criteria used in evaluation, and this disparity derives from their a periodicity and variable coverage. In this review, we identify three groups of studies: explicit evaluations, experimental evaluations and, more recently, several proposals for the establishment of a global framework to evaluate these systems.*

---

**Find other papers on this subject.**

---

### How to cite this paper:

Martínez Méndez, F. J. & Rodríguez Muñoz, J.V. (2003) "Síntesis y crítica de las evaluaciones de la efectividad de los motores de búsqueda en la web." *Information Research*, 8(2), paper no. 148 [Available at <http://InformationR.net/ir/8-2/paper148.html>]  
© the authors, 2003.  
Updated: (20 December 2002)

---

Check for citations, [using Google Scholar](#)

---

[Contents](#)

**17309**  
[Web Counter](#)

[Home](#)

---