

# Construcción de estrategias sistemáticas para la búsqueda exhaustiva de información en Internet: un marco de toma de decisiones aplicado a la información sobre psicología de la salud

[Julio Meneses](#), [Mercè Boixadós](#), [Lourdes Valiente](#), [Pep Vivas](#) y [Manuel Armayones](#)

Universitat Oberta de Catalunya

Estudis de Psicologia i Ciències de l'Educació.

Avda. del Tibidabo, 39-43. 08035 Barcelona (Catalunya), España.

---

## Abstract

**Background:** Con el aumento progresivo de los contenidos públicamente disponibles en Internet y a medida que la búsqueda de información se convierte en una de las actividades online más frecuente, la búsqueda informal y cotidiana de información no puede responder a los objetivos de la investigación científica. En la consideración de los contenidos relevantes para un determinado ámbito de conocimiento en un entorno no organizado como es Internet, surge la necesidad de crear procedimientos eficaces que garanticen una búsqueda sistemática.

**Objetivo:** Proponer un marco de referencia para la construcción de estrategias sistemáticas para la búsqueda exhaustiva en Internet, adaptable a los objetivos de búsqueda particulares.

**Método:** Revisión y propuesta teórica sobre el proceso de toma de decisiones en torno a la selección del método, instrumentos y la construcción de las ecuaciones de búsqueda. Ejemplificamos todo el proceso con la construcción de una estrategia sistemática en el ámbito de la información sobre Psicología de la Salud.

**Conclusiones:** La construcción de estos procedimientos o estrategias, articulados a partir de las propiedades de exhaustividad en la búsqueda, replicabilidad del procedimiento y validez ecológica, no sólo no es arbitraria, sino que está directamente relacionada con el tiempo empleado para la ejecución de la búsqueda y la calidad de la información obtenida.

---

## [Abstract in English](#)

---

## Breve contextualización de la búsqueda de información en Internet

Con el avance y difusión de las nuevas tecnologías y la generalización del acceso de la población a Internet, al finalizar el año 2004, el número de personas conectadas en el mundo alcanzó la cifra de aproximadamente 935 millones según los diferentes estudios de estimación de uso ([Clickz Stats 2004](#)), con una proyección que cifra la superación de los mil millones a principios del 2005 ([Computer Industry Almanac Inc. 2004](#)). Esto supone un incremento de en torno al 38% con respecto a los últimos datos ofrecidos por la [UNCTAD](#) (2004) para el 2003, un crecimiento interanual muy superior detectado en años anteriores que oscila entre un 7 y 27%. Paralelamente, y junto a la popularización del uso de este medio, cada vez más la búsqueda de información puntual a través de un buscador se está convirtiendo en una actividad frecuente (30%) tal y como indican los últimos informes de uso de Internet ([Pew Internet & American Life Project Tracking 2005](#)). Junto a ella, otras actividades que inspiran o

complementan la búsqueda de información ocupan también las posiciones más altas en las tablas de frecuencia de actividad en un día cualquiera, como por ejemplo la búsqueda de información de actualidad (27%), de la predicción del tiempo (22%), o de información para responder a una pregunta concreta (21%), solamente superadas globalmente por el uso del correo electrónico (48%). Además, no solamente ha crecido la población de usuarios de Internet y la búsqueda de información a través de ella, sino que además cada vez es mayor la cantidad de información disponible para hacerlo.

Dado el carácter anárquico y autorregulado de la red ([Abbate 2000](#)) no existe ningún censo o institución que ofrezca una medida fiable sobre la cantidad de sitios web o el número de documentos disponibles. Sin embargo, una medida indirecta como el número de hosts ofrecida por el ISC (Internet Software Consortium) puede servirnos para ilustrar su crecimiento. Este indicador hace referencia al número de dominios que tienen asignada una IP, y se estima conectando periódicamente con una muestra aleatoria sobre el número máximo teórico de IPs posibles con el actual sistema de asignación, tratando de verificar la existencia de un website. Así, el número de hosts estimados por el ISC ha pasado de 109.574.429 en enero de 2001 a 233.101.481 en enero de 2004, lo que supone un incremento de algo más del 112% en el número total de websites disponibles en Internet ([Internet Software Consortium 2005](#)). Sin embargo, dado que nuestra unidad de búsqueda no es el host o website sino los documentos que lo componen, debemos complementar esta medida indirecta de crecimiento con el de número de páginas disponibles dentro de un determinado ámbito de conocimiento. Tal como indica [Romero](#) (2002), se está produciendo un amplio desarrollo en el ámbito de la psicología gracias a la enorme y creciente cantidad de información y documentación públicamente disponible. Comparando los resultados obtenidos a partir de dos búsquedas informales y simbólicas en Google (google.com) utilizando como ecuación de búsqueda *psychology* por un lado y *psicologia* por otro con las que nos ofrece [Martín](#) (2001), el aumento en el número de documentos disponibles queda patente. En los cuatro años aproximadamente que han pasado entre ambas búsquedas se ha multiplicado entre siete y diez veces la información disponible en la red, pasando de 4.640.000 referencias que contenían el término *psychology* a 35.100.000, y de 190.000 a algo más de 2.070.000 con el término *psicologia*.

Desafortunadamente, determinar la información y los recursos disponibles en Internet en una determinada área de conocimiento, como paso previo a la evaluación de su calidad, no es un proceso tan sencillo. Debemos avanzar unos pasos más allá de las búsquedas informales basadas en las ocurrencias de términos, explotando todas las posibilidades que ofrecen los instrumentos de búsqueda ([ver definición](#)) en Internet, acercándonos así a los objetivos de la investigación científica. Se hace necesaria la utilización de una estrategia ([ver definición](#)) que afronte con las mayores garantías posibles la búsqueda sistemática de información en Internet.

## Requisitos para una búsqueda sistemática de información en Internet

El carácter flexible y descentralizado de la red al que aludíamos más arriba, virtud para la democratización del medio pero a la vez una desventaja para la efectividad de la búsqueda de información, requiere una reflexión profunda antes de comenzar a planificar una estrategia y así poder asegurar con unas ciertas garantías el haber sido lo suficientemente sistemáticos. Nos centramos en tres requisitos fundamentales, a saber: exhaustividad en la búsqueda, replicabilidad del procedimiento y validez ecológica.

### Exhaustividad en la búsqueda

El primer requisito, la exhaustividad en la búsqueda, afecta más al contenido que a la exactitud de la propia búsqueda, y es la condición que permite convenir que se han agotado todas las fuentes de información posibles dentro del universo de búsqueda ([ver definición](#)) que nos ocupa. La búsqueda de información en entornos estructurados como el catálogo de una biblioteca, un registro de publicaciones o una base de datos, parte de un universo de contenidos delimitado y cognoscible así como de un protocolo de búsqueda estandarizado. Sin embargo, en el caso de las búsquedas en la red no podemos hablar de estas condiciones de partida. Así, en una búsqueda en la base de datos PROQUEST, es posible conocer qué revistas y desde qué años, así como si se obtendrá la referencia, el abstract o el contenido completo de los artículos utilizando un procedimiento ya delimitado. Sin embargo, cuando damos el salto a Internet, ni conocemos el total de los potenciales sitios web en los que se hará la búsqueda, ni disponemos de un sistema de búsqueda pautado. Más aún, no sólo no conocemos la cantidad total de sitios en los que estamos realizando una búsqueda, sino que además se estima que solamente el 16% del total de contenidos realmente disponibles están indexados por los buscadores comerciales ([Lawrence y Giles 1999](#)). La velocidad con que la red evoluciona, crece, produce contenidos y renueva la información que ofrece es tan alta que es prácticamente imposible conocer, o siquiera hacerse una idea, de todos los contenidos

potencialmente disponibles en nuestro universo de búsqueda. Como afirman Oppenheim *et al.* ([2000](#)) el problema fundamental de Internet es el gran volumen de información, junto a la heterogeneidad e inconsistencia de ésta. 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. Si queremos establecer una estrategia eficaz de búsqueda de información en Internet, debemos tomar las precauciones necesarias para garantizar su exhaustividad y así no perder información relevante.

## Replicabilidad del procedimiento

En segundo lugar, la búsqueda de información en Internet debe garantizar la replicabilidad, es decir, debe ser posible realizar dos búsquedas simultáneas en las mismas condiciones y obtener los mismos resultados, así como establecer con cierta seguridad que las diferencias observadas en los resultados obtenidos en dos procesos de búsqueda independientes y separados en el tiempo se deban únicamente a la variación de los propios contenidos disponibles en Internet, y no por un déficit en la fiabilidad de la estrategia de búsqueda. Siguiendo el mismo ejemplo de búsqueda en una base de datos de artículos científicos como PROQUEST, donde el universo de búsqueda está altamente estructurado, el sistema de búsqueda que incorpora nos permite realizar búsquedas con una alta fiabilidad, de manera que cualquier variación en los resultados puede ser explicada única y exclusivamente por la incorporación de nuevos documentos al universo de búsqueda total.

## Validez ecológica

Y finalmente, en tercer lugar, la búsqueda debe asegurar una cierta validez ecológica, de manera que podamos concluir que se ajusta lo más posible a las condiciones que la elicitán. En muchos de los casos, sobre todo cuando tratamos de conocer la información públicamente disponible en un determinado ámbito de conocimiento, la estrategia de búsqueda deberá aproximarse lo más posible a la que cualquier usuario medio pueda llevar a cabo. Es en este sentido, en la medida en que las condiciones de registro reproducen con la máxima fidelidad aquellas en las que los fenómenos estudiados ocurren naturalmente ([León & Montero 2003](#)), en que podemos hablar de la necesidad de cumplir este requisito para establecer una estrategia de búsqueda sistemática, y donde aquellos contenidos no disponibles en la red, bien sea por no haber sido indexados por los buscadores, por estar únicamente disponibles para grupos privados, por problemas técnicos de disponibilidad, etc. ([Notess 2002](#); [Lawrence & Giles 1999](#)), no serán accesibles y por tanto irrelevantes para nuestros objetivos. A este respecto, debemos tener también presente que, aquellos recursos que dan acceso a lo que Ellsworth ([García 1996](#)) bautizó como Internet invisible (Invisible Internet) o Web profunda (Deep Web), tampoco sería interesante desde el punto de vista de la validez ecológica a pesar de que se estima que puede representar entre 400 y 550 veces más de información de que disponemos en la parte *pública* ([Bergman 2000](#)).

Tratando de responder a estos tres requisitos, nuestra propuesta de estrategia de búsqueda de información en Internet parte del análisis de los diferentes métodos de búsqueda ([ver definición](#)) utilizados convencionalmente por la población general ([Barker 2003](#); [Notess 2002](#); [The University of Texas System Digital Library 2003](#)).

## Métodos de búsqueda de información en Internet

Aunque a lo largo del artículo hemos ido definiendo cada uno de los conceptos claves en la búsqueda de información, es necesario hacer una precisión terminológica antes de revisar los diferentes métodos de búsqueda disponibles en la red. Si bien en el lenguaje común las expresiones *método de búsqueda* y *estrategia de búsqueda* suelen utilizarse de forma equivalente, es conveniente precisar que no se trata de la misma cosa. Mientras el método es una perspectiva adoptada antes de plantear una búsqueda de información concreta, la estrategia consiste en la serie de pasos que nos permite estructurar la propia búsqueda. No todo método de búsqueda implica el uso de una estrategia, pero toda estrategia debe considerar al menos un método de búsqueda.

Así, partiendo de las clasificaciones al uso ([Barker 2003](#); [University of Texas 2003](#); [Notess 2002](#); [Sullivan 2002b](#)) es posible distinguir fundamentalmente cinco métodos de búsqueda de información en Internet cuya lógica describiremos brevemente a continuación:

### El uso de directorios temáticos de sitios web

Un directorio temático es un instrumento de búsqueda de información dirigido por un equipo de editores que valoran el alta o modificación de las solicitudes de inclusión en el directorio. Disponen, como otros instrumentos de búsqueda de información en Internet, de un sistema de entrada y de una base de datos que generalmente contiene el nombre del recurso, las palabras clave que lo identifican y una descripción que los responsables del recurso adjuntan en su solicitud. La actualización de su base de datos, al contrario que los buscadores automatizados que veremos a continuación, se nutre de las solicitudes enviadas por los responsables de los recursos, de manera que la velocidad y precisión con que refleja el universo total de recursos disponibles en Internet son relativas. Aunque en los primeros tiempos de la popularización de Internet fue uno de los sistemas de organización de información más importantes comandado por el paradigmático Yahoo! ([yahoo.com](http://yahoo.com)), actualmente están muy en desuso por las limitaciones de su sistema de actualización.

## La consulta en bases de datos especializadas

Las bases de datos especializadas son un servicio de búsqueda generalmente dedicado a la consulta de artículos en revistas científicas, fondos bibliotecarios, etc., y no necesariamente orientadas a la catalogación de sitios web. Las bases de datos tradicionales han encontrado una vía de explotación ubicua a través de Internet, pero se trata de un recurso especializado generalmente poco utilizado por la población general. Además, no suelen estar indexadas por los buscadores y directorios más generales, formando parte de lo que llamábamos más arriba la Internet invisible.

## El uso de buscadores automatizados

Un buscador automatizado es un instrumento de búsqueda de información donde la inclusión y la modificación de los recursos disponibles en Internet se realiza automáticamente gracias al uso de robots o arañas, complementando las fuentes de entrada con la sugerencia por parte de terceros al estilo de los directorios temáticos. Un robot o araña (spider o bot en inglés) es una aplicación que simula la actividad de un usuario navegando en Internet, a través del cual los buscadores automatizados actualizan su base de datos sobre los recursos disponibles. A partir de una página conocida, la araña o robot rastrea de forma cíclica y sistemática los enlaces ofrecidos, saltando de recurso en recurso, y accediendo a su contenido para realizar la copia literal local. Si durante el proceso la araña descubre un enlace a un recurso no contemplado hasta el momento en la base de datos, o el contenido disponible en alguna de las páginas de una fuente ya indexada ha variado con respecto a la última visita, esta información será añadida a la base de datos de manera que la búsqueda se haga sobre una representación lo más exacta posible de los contenidos realmente disponibles en Internet.

Como en el caso de los directorios temáticos, disponen de un sistema de entrada y de una base de datos, pero en este caso contiene el nombre del recurso y una copia literal local de los contenidos originales ofrecidos por sus responsables. La búsqueda, al contrario que en los directorios temáticos y gracias al avance de las técnicas de recogida y almacenamiento de información, no se limita a la coincidencia de los términos de la ecuación de búsqueda con el nombre del recurso o la descripción proporcionada, sino que se amplía al contenido de cada una de las páginas indexadas (Barroso *et al.* 2003; Brin & Page 1998). Los buscadores automatizados, por la potencia que les otorga este tipo de representación de los contenidos disponibles en Internet, se han convertido en un recurso clave con Google.com a la cabeza. Este instrumento de búsqueda se ha convertido en los últimos años en el buscador automatizado por excelencia al haber podido hacer realidad un sistema de búsqueda autoactualizable de forma indefinida únicamente limitado por la capacidad máxima teórica del hardware utilizado para dar soporte a la información (Notess 2003; Brin & Page 1998).

## La solicitud de ayuda a un experto

Las redes sociales personales son una fuente imprescindible de información (Johnson 2004), y recurrir a un experto cuando se necesita ayuda es una de las formas más cómodas y evidentes de conseguir una determinada información, donde se hace valer una máxima de nuestro tiempo donde *lo importante no es saber, sino tener el e-mail del que sabe*. Como veremos más adelante, una interpretación no tan literal de este método podría resultar interesante para desarrollar una estrategia de búsqueda sistemática.

## La suerte

Cajón de sastre en el que se agrupan las búsquedas no convencionales, destacando por ejemplo la ingenua creencia

de que se puede encontrar cualquier recurso en Internet simplemente escribiendo su nombre con estructura de sitio web entre [www.](#) y [.com](#). La suerte, junto a la solicitud de ayuda a un experto, son los dos métodos de búsqueda menos sistematizados gracias a su escasa estructuración, aunque no por ello resultan ser los menos utilizados. La incorrecta pero a la vez popular creencia de que el recurso, probablemente el de mayor calidad, ha de tener como dirección su nombre con estructura de sitio web entre [www.](#) y [.com](#) es tan frecuente que son innumerables los casos de compleja disputa por su posesión. Un ejemplo muy significativo es el de [sex.com](#), un negocio redondo en publicidad rentabilizando este método de búsqueda sobre una de las temáticas estrella en la búsqueda a través de la red ([Glasner 2003](#)).

A partir del análisis de las ventajas e inconvenientes de las diferentes alternativas, propondremos una estrategia sistemática y exhaustiva que abarque un universo lo más extenso posible para una población determinada, adaptable a las necesidades particulares de cada búsqueda.

## Construcción de una estrategia de búsqueda sistemática

En la segunda parte de nuestro artículo, revisados los conceptos fundamentales necesarios para la construcción de una estrategia de búsqueda sistemática, guiaremos al lector interesado en este tipo de búsquedas a través de un proceso de toma de decisiones en torno a la selección del método de búsqueda y de los instrumentos, la construcción de las ecuaciones de búsqueda ([ver definición](#)) que nos permita sistematizar una estrategia adaptándola a los objetivos de búsqueda particulares. Cualquier decisión adoptada a partir de las recomendaciones propuestas en el presente artículo deberá estar siempre dirigida por el principio de validez ecológica, es decir, tratando siempre como referencia las condiciones para las que construimos la estrategia sistemática. Finalmente, ejemplificaremos nuestro discurso en la aplicación desarrollada para el proyecto 'La Psicología de Salud y la Calidad de Vida en la Sociedad de la Información y el Conocimiento' (IN3 IR220), dirigido a la evaluación de la calidad de los contenidos en Internet desde el punto de vista de la Psicología de la Salud ([Vivas, Armayones, Boixadós, Herrero, Meneses, Suelves & Valiente 2003](#)).

### Selección del método de búsqueda

Dado que la búsqueda informal de información en Internet se asocia al uso de un buscador con una ecuación de búsqueda improvisada, no se suele distinguir con precisión entre los diferentes métodos de búsqueda disponibles. Debido a la facilidad y rapidez con que se pueden obtener resultados relevantes para este tipo de búsquedas, no hay un interés explícito por reflexionar sobre las diferentes alternativas disponibles. Sin embargo, una búsqueda sistemática no debe pasar por alto este aspecto. Decidir qué método es el que mejor se ajusta a nuestras necesidades de búsqueda es, en otras palabras, decidir qué tipo de información y cuánta estamos dispuestos a despreciar en función de la relación calidad de la información obtenida/tiempo empleado para la búsqueda. Disponer de unas nociones mínimas acerca de los métodos como hemos pretendido más arriba, es la clave para poder tomar las decisiones oportunas según nuestro objetivo de búsqueda.

Descartadas la búsqueda en bases de datos especializadas por no ser el estudio de la parte invisible u oculta de Internet nuestro objetivo, y la estrategia basada en la suerte por no ser un método estructurado y sistemático, excluirémos la búsqueda a través de directorios temáticos por su importante limitación en la actualización de sus contenidos con respecto a los buscadores automatizados ([Brin y Page 1998](#); [Sullivan 2002b](#)). A priori, teniendo en cuenta los requisitos necesarios para una búsqueda sistemática, podríamos considerar a los buscadores automatizados como el método idóneo por su mayor exactitud en la representación de los contenidos realmente disponibles en Internet. Sin embargo, al no tratarse de un sistema supervisado, debemos considerar también sus limitaciones no sólo para tener en cuenta el tipo de información que podríamos perder en la búsqueda, sino para tratar de compensarlo en nuestra estrategia.

Por un lado, la automatización del proceso a partir de la coincidencia de términos requiere del juicio de experto para la desestimación de recursos irrelevantes (falsos positivos). A medida que los contenidos indexados por sus bases de datos crece, y en espera del desarrollo y perfeccionamiento de herramientas que permitan al propio buscador ponderar la importancia de los resultados obtenidos en función de los objetivos de búsqueda, la gran cantidad de falsos positivos que devuelve una única ecuación de búsqueda en un único buscador hace imposible revisarlas todas una a una ([University of Texas 2003](#)). En el ejemplo de búsqueda a través de Google ([google.com](#)) que presentamos al inicio de este artículo, el número de referencias que contenían el término *psychology* es de 8.010.000. Como se puede adivinar con un sencillo cálculo, suponiendo que una persona invirtiera únicamente un segundo en consultar



cada una de las referencias, tardaría algo más de 92 días (leyendo las 24 horas del día) en consultar todas ellas y así determinar si resultan relevantes o no para los objetivos de búsqueda. Incluso para aquellos que puedan leer el contenido de cada uno de los sitios web en algo menos de un segundo, 92 días de lectura continua resultan algo difícilmente manejable. Más aún, siendo críticos con el propio sistema de coincidencias en el que se basan, no siempre un resultado positivo lleva aparejada una respuesta correcta ([Sullivan 2002a](#)). Por otro lado, ya que su base de datos depende de la propagación de las arañas o robots que conforman su sistema de actualización, debemos tener presente que los recursos obtenidos como relevantes podrían no ser todos los que están realmente disponibles (falsos negativos). Basar nuestra estrategia en el uso exclusivo de este método de búsqueda podría comportar no considerar como relevantes aquellos recursos que simplemente no han sido indexados en el proceso de actualización ([Notess 2002](#)).

Así pues, recogiendo los puntos críticos de los buscadores automatizados que acabamos de revisar, en la selección del método de búsqueda trataremos de reducir los falsos negativos y falsos positivos, así como el sesgo que supone la utilización exclusiva de métodos de búsqueda automatizados. En primer lugar, reduciremos los falsos negativos ampliando el número de buscadores en los que poner en práctica nuestra estrategia. Limitar a un único instrumento de búsqueda tiene la ventaja de evitar los solapamientos que pudiera haber en la consideración de los recursos potencialmente relevantes, pero como hemos visto puede ser a la vez una importante fuente de sesgo para una búsqueda sistemática. A pesar de que todos ellos comparten la lógica de expansión y actualización descrita más arriba, la aplicación concreta de cada uno de ellos a partir de unas condiciones diferenciales (recursos de partida, potencia de los robots o arañas, condiciones técnicas que determinan el tamaño de las bases de datos, etc.) introduce la variación entre sus bases de datos y por tanto la ausencia de solapamiento. Si este solapamiento no es completo, es decir, la representación de los recursos realmente disponibles difiere entre los diferentes buscadores, la utilización de una misma ecuación en un momento concreto no conlleva necesariamente la obtención de idénticos resultados. Reducir los falsos negativos es entender la codificación y almacenamiento de los recursos disponibles en Internet como una característica variable, y por lo tanto apostar por no limitar la búsqueda sistemática a un único instrumento de búsqueda.

En segundo lugar, reduciremos los falsos positivos no sólo acudiendo al juicio de experto para evaluar la pertinencia de cada uno de ellos en función de los objetivos de búsqueda, sino reduciendo el número de recursos que deben ser evaluados por cada ecuación de búsqueda. Para evitar la información no relevante o redundante, es habitual recomendar la reducción del número de referencias a considerar a partir de una ecuación dada. En el balance entre la calidad de la información obtenida y el tiempo exigido para ella, consideramos razonable moverse entre los diez primeros resultados que suelen ser habituales en una búsqueda más informal ([Madrid & Gauch 2003](#)), y una postura más conservadora en torno a los veinte o treinta primeros que nos asegurarán una mayor confianza en la exhaustividad de nuestra búsqueda.

Y por último, en tercer lugar, reduciremos el sesgo que supone la utilización exclusiva de métodos de búsqueda automatizados considerando la combinación con una revisión de lo que más arriba llamamos la petición de ayuda a un experto. Si en la búsqueda de información para unos objetivos concretos encontramos un recurso relevante, la petición de ayuda al experto podría entenderse como la consideración de los recursos propuestos por los responsables del propio recurso en la habitual sección de enlaces. Aunque se trata de un método no automatizado y, por lo tanto, potencialmente sujeto a las mismas críticas que los directorios temáticos, la especialización esperable por parte de sus responsables en su área de conocimiento, nos servirá para evaluar otros que, siendo potencialmente pertinentes para nuestros objetivos de búsqueda, bien podrían quedar fuera del alcance de los buscadores automatizados ([Meneses et al. 2003](#)). Si la falta de especialización por parte de los editores a la hora de construir y mantener los directorios temáticos puede explicar una buena parte de sus limitaciones en la representación de los contenidos realmente disponibles en Internet, en esta variante de ayuda al experto encontraríamos una buena solución para evitar el sesgo derivado de la utilización en exclusiva de buscadores automatizados.

## **Selección de los instrumentos de búsqueda**

Una vez establecido el método, la elección de los instrumentos de búsqueda no es una cuestión baladí. En el momento en que nos encontramos en la construcción de una estrategia sistemática, la noción de validez ecológica que introdujimos más arriba debe servirnos de guía. Lejos de ser una cuestión de preferencias o suposiciones, la elección de los buscadores que servirán de instrumentos de búsqueda responderá a un trabajo serio y riguroso en busca de los más adecuados para nuestros objetivos particulares. Así, en función de la temática y la población que nos interesen, será necesaria la revisión de los diferentes informes de las principales auditoras del uso de Internet

que publican regularmente empresas como comScore Media Metrix ([comscore.com](http://comscore.com)), Cooperative Association for Internet Data Analysis ([caida.org](http://caida.org)), Gartner ([gartner.com](http://gartner.com)), iProspect ([iprospect.com](http://iprospect.com)), Jupiter Research ([jupiterresearch.com](http://jupiterresearch.com)), Nielsen//Netratings ([nielsen-netratings.com](http://nielsen-netratings.com)) o Search123 ([search123.com](http://search123.com)) entre otros. Esta colección de auditoras, lejos de ser exhaustiva e invariable, se ofrece como pequeña orientación válida para el momento en que se editan estas líneas. Además, en el caso de que los objetivos de búsqueda estén relacionados con una población muy concreta como ocurre en el caso que introduciremos más adelante como ilustración, los investigadores deberán considerar también aquellos publicados por organismos y entidades de ámbito local.

## Construcción de las ecuaciones de búsqueda

A pesar de que aquí seguiremos la estructura de las ecuaciones de búsqueda más habituales, es importante dedicar un pequeño esfuerzo al estudio de la sintaxis de los buscadores elegidos para así optimizar el rendimiento de nuestras búsquedas. Una revisión comparativa de las principales formas en [Barker](#) (2003), aunque siempre es recomendable la consulta de la documentación de los propios responsables del instrumento.

Las ecuaciones de búsqueda son, a grandes rasgos, el medio a través del cual interactuamos con el sistema de entrada de los buscadores. Aunque su sintaxis puede variar de un instrumento a otro, en líneas generales podemos recomendar una serie de directrices a la hora de su planificación y construcción. En primer lugar, entendiendo los términos como el léxico para la comunicación con el buscador, debemos hacer una selección lo suficientemente extensa como para agotar la variabilidad que puede haber en el objetivo inicial de búsqueda. Es importante tener en cuenta que al tratarse de instrumentos de búsqueda que realizan una copia literal de los contenidos en su base de datos, los términos elegidos aumentarán su efectividad más que por ser fruto de una elaboración puramente conceptual o teórica, por ser frecuentemente utilizados en los recursos que nos interesan. En un entorno de organización de la información basado en la ocurrencia de términos, la búsqueda se convierte en el sondeo de la red, siguiendo la pista de los documentos que contienen las palabras que conforman las ecuaciones.

Una vez elegidos los términos, construiremos las ecuaciones de búsqueda utilizando la sintaxis oportuna para cada uno de los instrumentos de búsqueda. Dada la flexibilidad actual de los buscadores automatizados, por norma general será suficiente utilizar los términos diana, combinando si fuera necesario los diferentes elementos sin operadores, en minúsculas y sin tildes, utilizando las comillas únicamente para la búsqueda de expresiones concretas. Aunque en los primeros pasos en el almacenamiento y recuperación de información en Internet se aplicaron criterios de búsqueda de la documentación tradicional, la popularización de los buscadores automatizados ha llevado a su simplificación por ejemplo al asumir la utilización del operador AND al introducir dos términos separados por un espacio. En cuanto al uso de comillas para buscar expresiones, se ha de tener en cuenta que la ocurrencia de ambos términos no relacionados en el mismo documento podría aumentar los falsos positivos, promoviendo la consideración de recursos que pueden no ser relevantes para los objetivos de búsqueda. Teniendo en cuenta que no se considera la totalidad de referencias devueltas por un buscador, el ruido introducido entre los resultados de los buscadores produciría la pérdida de información potencialmente relevante. En un ejemplo de búsqueda sobre trastornos musculares, no utilizar comillas en la ecuación *codo de tenista* sin duda aumentará el ruido en la búsqueda de referencias relevantes para nuestros objetivos.

## Sistematización de la estrategia de búsqueda: un ejemplo en Psicología de la Salud

A continuación y para finalizar, ilustraremos el proceso de construcción de una estrategia sistemática que estamos describiendo con el que llevamos a cabo en el marco del proyecto 'La Psicología de Salut i la Qualitat de Vida en la Societat de la Informació i el Coneixement' ([Vivas et al. 2003](#)). El proyecto, uno de los pioneros en nuestro país en el análisis de calidad de la información sobre salud disponible en Internet, dirigió nuestros intereses de búsqueda en la información públicamente accesible en Internet, tanto en catalán como en español, relacionada con la Psicología de la Salud. La búsqueda fue conducida por nuestros investigadores entre julio y noviembre de 2002, y los resultados almacenados en un directorio privado online para después proceder a la copia local de los recursos y posterior análisis de contenido. Así, utilizando los principios para una estrategia de búsqueda sistemática que hemos presentado, seleccionamos el método de búsqueda, los instrumentos y construimos las ecuaciones de búsqueda pertinentes.

A partir de una consideración minuciosa de los estudios de audiencias de servicios en Internet para nuestra población

(Nielsen//Netratings 2002; Asociación para la Investigación de los Medios de Comunicación 2001), seleccionamos los cinco buscadores automatizados más utilizados, a saber: Terra ([terra.es](http://terra.es)), Altavista ([altavista.com](http://altavista.com)), Google ([google.com](http://google.com)), Lycos ([lycos.es](http://lycos.es)) y MSN ([msn.es](http://msn.es)). A pesar de la popularidad informada del buscador automatizado de Yahoo! Inc. ([yahoo.es](http://yahoo.es)) en los estudios de referencia, su inclusión entre los instrumentos de búsqueda pertinentes fue rechazada al tener su servicio subcontratado a Google en el periodo en que planificamos y realizamos la búsqueda (Sullivan 2002c). A continuación, construimos las ecuaciones de búsqueda a partir de la combinación de los términos referidos a los trastornos con relevancia para la Psicología de la Salud y los diferentes niveles de intervención, seleccionados a partir de la revisión de la literatura al uso. En el caso de los trastornos, fueron elegidos aquellos relevantes para la Psicología de la Salud a partir del listado de enfermedades descritas por la OMS (Organización Mundial de la Salud, n.d.), y en el caso de los niveles de intervención los términos fueron elegidos a partir de la revisión de los manuales al uso de la disciplina (León & Medina 2002; Brannon & Feist 2001; Amigo *et al.* 1998). Siguiendo las indicaciones para la construcción de ecuaciones que presentábamos más arriba, introdujimos además en todas las ecuaciones el término *psicología* para reducir los falsos positivos y así aumentar la probabilidad en la detección de recursos relevantes para la disciplina.

## Formulación de la estrategia de búsqueda

Una vez aquí, y tras un ensayo realizado por los propios investigadores para demostrar la necesidad de utilizar una combinación de los métodos de búsqueda automatizada y ayuda de experto (Meneses *et al.* 2002), pusimos en marcha la búsqueda a través de los pasos que siguen ([ver cuadro 1](#)):

### 1. Acceso vía web al primero de los cinco buscadores elegidos

A partir de los buscadores automatizados seleccionados para este estudio (buscador #1, buscador #2, buscador #3, buscador #4 y buscador #5), accedimos a la web del primero para realizar la búsqueda.

### 2. Introducción de las ecuaciones de búsqueda en los sistemas de entrada

Construidas las trescientas diez ecuaciones de búsqueda resultantes de la combinación de los treinta y un trastornos, los diez niveles de intervención seleccionados y el término "psicología", introdujimos la primera en el sistema de entrada del buscador para enviar la consulta al servidor.

### 3. Selección del primer recurso relevante de entre los resultados obtenidos en la búsqueda e inclusión en el buscador

Obtenida la página de resultados para la ecuación de búsqueda, cada una de las treinta primeras referencias propuestas por el buscador fueron visitadas para seleccionar la primera relevante de acuerdo a los objetivos de búsqueda. Se considera una referencia relevante cuando aborda alguna de las áreas de interés, excluyendo en todo caso recursos puntuales no pertinentes, como son las webs de bibliotecas o catálogos bibliográficos, los diarios y los programas de asignaturas universitarias. Estos recursos fueron manejados como falsos positivos, al contener las palabras claves a partir de las que realizamos la busca pero resultar irrelevantes para los objetivos de búsqueda.

### 4. Consideración de los recursos propuestos por los responsables

Una vez aceptado el recurso localizado a través del buscador automatizado, revisamos su contenido en busca de otras referencias propuestas como relevantes en su ámbito, generalmente agrupadas en un apartado llamado enlaces, directorio de recursos, links de interés, etc. Localizado el apartado, revisamos todas y cada una de las referencias propuestas, tratándolas de forma idéntica a las referencias ofrecidas por el buscador, incluyendo los recursos relevantes en el directorio y valorando de nuevo las referencias propuestas en cada uno de ellos.

### 5. Selección y exploración de los siguientes recursos relevantes de entre los resultados ofrecidos por el buscador automatizado

Evaluada los enlaces propuestos por los responsables del primer recurso seleccionado anteriormente de entre los resultados del buscador, continuamos nuestra búsqueda repitiendo el proceso desde el paso 3 de forma cíclica hasta



llegar a la referencia número treinta.

## **6. Acceso vía web y exploración de los resultados ofrecidos por cada uno de los restantes buscadores automatizados**

A partir de este momento, el proceso de búsqueda comenzó de nuevo, manteniendo la misma ecuación de búsqueda pero variando el buscador, eligiendo los recursos relevantes así como los propuestos por estos tal y como se detalla en los pasos anteriores.

## **7. Repetición de la anterior secuencia utilizando cada una de las diferentes ecuaciones de búsqueda construidas**

Agotadas todas las combinaciones posibles de la ecuación de búsqueda con cada uno de los buscadores escogidos, el proceso comenzó de nuevo con la siguiente ecuación hasta agotar a las algo más de trescientas.

De esta forma, llevamos a cabo un total de 1550 búsquedas (310 ecuaciones de búsqueda x 5 buscadores), evaluando la pertinencia de un total de 46.500 referencias (1550 búsquedas x 30 referencias evaluadas en cada una), así como las propuestas de los responsables de cada uno de los recursos relevantes para los objetivos de búsqueda. Esta estrategia sistemática nos permitió la localización de 368 recursos relevantes para la Psicología de la Salud 23 en catalán y 345 en castellano. Los resultados de esta investigación, así como los detalles en la ejecución de la búsqueda, pueden leerse en [Vivas et al.](#) (2003).

## **Conclusiones**

Las principales conclusiones que se derivan de este estudio se presentan de forma esquemática a continuación.

- Con la progresiva difusión de Internet como medio para la distribución y acceso a la información, surge la necesidad de crear sistemas eficaces para la búsqueda sistemática de información relevante, en busca de las propiedades de exhaustividad, replicabilidad del procedimiento y validez ecológica. El contexto de búsqueda aquí expuesto, así como la estrategia que hemos presentado como ejemplo, pretenden servir de marco de referencia en cualquier área de conocimiento.
- No es posible, además de ser algo arriesgado, asegurar que una determinada estrategia de búsqueda será lo suficientemente exhaustiva como para no perder ningún recurso relevante. La exhaustividad, lejos de ser una cualidad de todo o nada, es una dimensión íntimamente relacionada con la relación entre la calidad de la información obtenida y el tiempo empleado para la búsqueda.
- La exploración sistemática de los contenidos ofrecidos en Internet, independientemente de los objetivos a los que responda, requiere de una reflexión más profunda que la que se da en el caso de las búsquedas informales. Como hemos demostrado, la elección de un método y no otro, la población para la que realizamos la búsqueda, los instrumentos que elegimos y la propia estrategia de búsqueda con que operativizamos el proceso, no sólo no son decisiones arbitrarias sino que determinan en un alto grado los resultados obtenidos.
- Al contemplar la información disponible en la red como una parte importante de la información públicamente disponible en las diferentes áreas de conocimiento, consideramos a los responsables de los propios recursos como expertos en el área. Sin abordar cuestiones relativas al dominio de los contenidos, tienen un papel fundamental en la selección de los enlaces de los recursos relevantes para su área de conocimiento. Considerarlos como una de las principales puertas de entrada a la información es, no sólo acercar la búsqueda sistemática a las pautas utilizadas generalmente en la navegación informal, sino evitar la limitación que supone una búsqueda articulada sobre la mera ocurrencia de términos.
- Al contrario de lo que podría suponerse a priori, el uso de buscadores automatizados puede no ser la mejor respuesta ante una necesidad de búsqueda sistemática. En todo caso, no es la única alternativa, ni recomendamos su uso en exclusiva a pesar de ser uno de los sistemas de búsqueda y recuperación de información más complejos y completos.
- El presente artículo además, pretende servir de ayuda en la precisión terminológica en el marco de las búsquedas de información en Internet. En muchas ocasiones, conceptos como el método o la estrategia de búsqueda son utilizados de forma indistinta, y no son muchos los lugares en los que se argumente claramente a favor y en contra de un determinado método desde el punto de vista de las garantías que puede ofrecer en la aproximación sistemática. Concretamente la propuesta aquí presentada, después analizar todas las ventajas e

inconvenientes se basa en la selección de los buscadores automatizados como el método más idóneo compensando sus limitaciones con nuestra estrategia sistemática de búsqueda de información en Internet.

- La evaluación de calidad de los contenidos públicamente disponibles en la red, y más aún en el área de la salud por las implicaciones que el uso de información incorrecta o de mala calidad puede suponer, no es más que el examen de los resultados obtenidos tras la ejecución de una estrategia de búsqueda.

## Agradecimientos

Esta investigación ha sido realizada, en parte, gracias a la ayuda del Internet Interdisciplinary Institute (IN3IR220); del Departament d'Universitats, Recerca i Societat de la Informació de la Generalitat de Catalunya (2003FI 00016); y del Ministerio de Ciencia y Tecnología (SEC2003-03403/PSCE).

## Referencias

- Abbate, J. (2000). *Inventing the Internet*. Cambridge: Mit Press.
- Amigo, I., Fernández, C. & Pérez, M. (1998). *Manual de psicología de la salud*. Madrid: Pirámide.
- Asociación para la Investigación de los Medios de Comunicación (2001). [Navegantes en la red. Cuarta encuesta AIMC a usuarios de Internet](#). Retrieved 20 December 2003 from AIMC Website, <http://www.aimc.es/03internet/macro2001.pdf>
- Barker, J. (2003). [Finding information on the Internet: a tutorial](#). Retrieved 20 December 2003 from University of California, Library Berkeley Web site: <http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/FindInfo.html>
- Barroso, L.A., Dean, J., Hölzle, U. (2003). Web search for a planet: the Google cluster architecture. *IEEE Micro*, **23**(2) 22-28.
- Bergman, M.K. (2000). [The deep Web: surfacing hidden value](#). Retrieved 20 December 2003 from Brighplanet Website, <http://www.brightplanet.com/deepcontent/>
- Brannon, L. & Feist, J. (2001). *Psicología de la salud*. Madrid: Paraninfo.
- Brin, S. & Page, L. (1998). The anatomy of a large scale hypertextual Web search engine. *Computer Networks and ISDN Systems*, **30**(1-7) 107-117.
- Clickz Stats (2004). [Population explosion!](#) Retrieved 1 January 2005 from Clickz website, [http://www.clickz.com/stats/sectors/geographics/article.php/5911\\_151151](http://www.clickz.com/stats/sectors/geographics/article.php/5911_151151)
- Computer Industry Almanac Inc. (2004). [Worldwide Internet users will top 1 billion in 2005](#). Retrieved 1 January 2005 from CIA Website, <http://www.c-i-a.com/pr0904.htm>.
- García, F. (1996, January 19). Business and marketing on the Internet. *The Computer Paper*, **9**(1).
- Glasner, J. (2003, January 4). [Court punts Sex.com domain case](#). *Wired News*. Retrieved 20 December 2003 from <http://www.wired.com/news/business/0,1367,57065,00.html>
- Internet Software Consortium (2005). [ISC Internet Domain Survey](#). Retrieved 1 January, 2005 from Internet Software Consortium Website, <http://www.isc.org/index.pl?/ops/ds/>
- Johnson, C. A. (2004). [Choosing people: the role of social capital in information seeking behaviour](#). Information Research, **10**(1), paper 201. Retrieved 1 January, 2005 from <http://informationr.net/ir/10-1/paper201.html>
- Lawrence, S. & Giles, C.L. (1999). [Accessibility and distribution of information on the Web](#). *Nature*, **400**(107), 107-109. Retrieved 18th April, 2005 from [http://www.nature.com/cgi-taf/DynaPage.taf?file=/nature/journal/v400/n6740/full/400107a0\\_fs.html](http://www.nature.com/cgi-taf/DynaPage.taf?file=/nature/journal/v400/n6740/full/400107a0_fs.html) (Subscription required.)
- León, J. M. & Medina, S. (2002). *Psicología social de la salud. Fundamentos teóricos y metodológicos*. Sevilla: Comunicación Social Ediciones y Publicaciones.
- León, O. G. & Montero, I. (2003). *Métodos de investigación en psicología y educación (3rd Ed.)*. Madrid: McGraw-Hill.
- Madrid, J. M. & Gauch, S. (2003). KeyConcept: un motor de búsqueda conceptual. *Sistemas & Telemática*, **1**(1), 47-62.
- Martín, J. (2001, December 17). Googlandia. *Ciberpaís*, 25-29.
- Meneses, J., Armayones, M., Boixadós, M., Suelves, J. M., Valiente, L., & Vivas, P. (2003). Búsqueda exhaustiva de información de Internet: una propuesta de protocolo desde la Psicología. *Encuentros en Psicología Social*, **1**(1) 133-138.
- Nielsen//NetRatings (2002). [Hispanics are the fastest growing ethnic group](#). Retrieved 20 December 2003 from Nielsen//Netratings Website, [http://www.nielsen-netratings.com/pr/pr\\_020716.pdf](http://www.nielsen-netratings.com/pr/pr_020716.pdf)

- Notess, G. (2002). [Learning about searching](http://www.searchengineshowdown.com/strat/). Retrieved 20 December 2003 from Search Engine Showdown Website, <http://www.searchengineshowdown.com/strat/>
- Notess, G. (2003). [Search engine statistics](http://www.searchengineshowdown.com/stats/). Retrieved 20 December 2003 from Search Engine Showdown Website, <http://www.searchengineshowdown.com/stats/>
- Organización Mundial de la Salud (n.d.). [Temas de salud](http://www.who.int/health_topics/es/). Retrieved 20 December 2003 from WHO Website, [http://www.who.int/health\\_topics/es/](http://www.who.int/health_topics/es/)
- Oppenheim, C., Morris, A., McKnight, C. & Lowley, S. (2000). The evaluation of WWW search engines. *Journal of Documentation*, **56**(2) 190-211.
- Pew Internet & American Life Project Tracking (2005). [Daily Internet activities. The percentage of Internet users who do a given activity on a 'typical day'](http://www.pewinternet.org/trends.asp). Retrieved 1 January, 2005 from PEW Website, <http://www.pewinternet.org/trends.asp>
- Romero, A. (2002). La nueva herramienta de información, documentación y comunicación: Internet. In F.M. Tortosa, C. Civera (Eds.), *Nuevas tecnologías de la información y documentación en psicología* (pp. 163-197). Barcelona: Ariel.
- Sullivan, D. (2002a). [An Egyptian stumper for Google, or is it?](http://www.searchenginewatch.com/searchday/article.php/2160001) Retrieved 20 December, 2003 from Search Engine Watch Website, <http://www.searchenginewatch.com/searchday/article.php/2160001>
- Sullivan, D. (2002b). [How Search Engines Work](http://www.searchenginewatch.com/webmasters/article.php/2168031). Retrieved 20 December, 2003 from Search Engine Watch Website, <http://www.searchenginewatch.com/webmasters/article.php/2168031>
- Sullivan, D. (2002c). [Yahoo Renews With Google. Changes Results](http://searchenginewatch.com/sereport/article.php/2165081). Retrieved 20 December, 2003 from Search Engine Watch Website, <http://searchenginewatch.com/sereport/article.php/2165081>
- The University of Texas System Digital Library (2003). [Texas information literacy tutorial](http://tilt.lib.utsystem.edu/). Retrieved 20 December, 2003 from University of Texas System Digital Library Website, <http://tilt.lib.utsystem.edu/>
- United Nations Conference on Trade and Development (2004). [E-commerce and development report 2004](http://www.unctad.org/en/docs/ecdr2004_en.pdf). Retrieved 1 January, 2005 from [http://www.unctad.org/en/docs/ecdr2004\\_en.pdf](http://www.unctad.org/en/docs/ecdr2004_en.pdf)
- Vivas, P., Armayones, M., Boixadós, M., Herrero, J. Meneses, J., Suelves, J. M. & Valiente, L. (2003). [Salut i Internet: cerca, temàtica i qualitat del seu contingut](http://www.uoc.edu/in3/psinet/docs/publicaciones/working02.pdf). Retrieved 20 December, 2003 from Universitat Oberta de Catalunya, PSiNET Group Web site, <http://www.uoc.edu/in3/psinet/docs/publicaciones/working02.pdf>

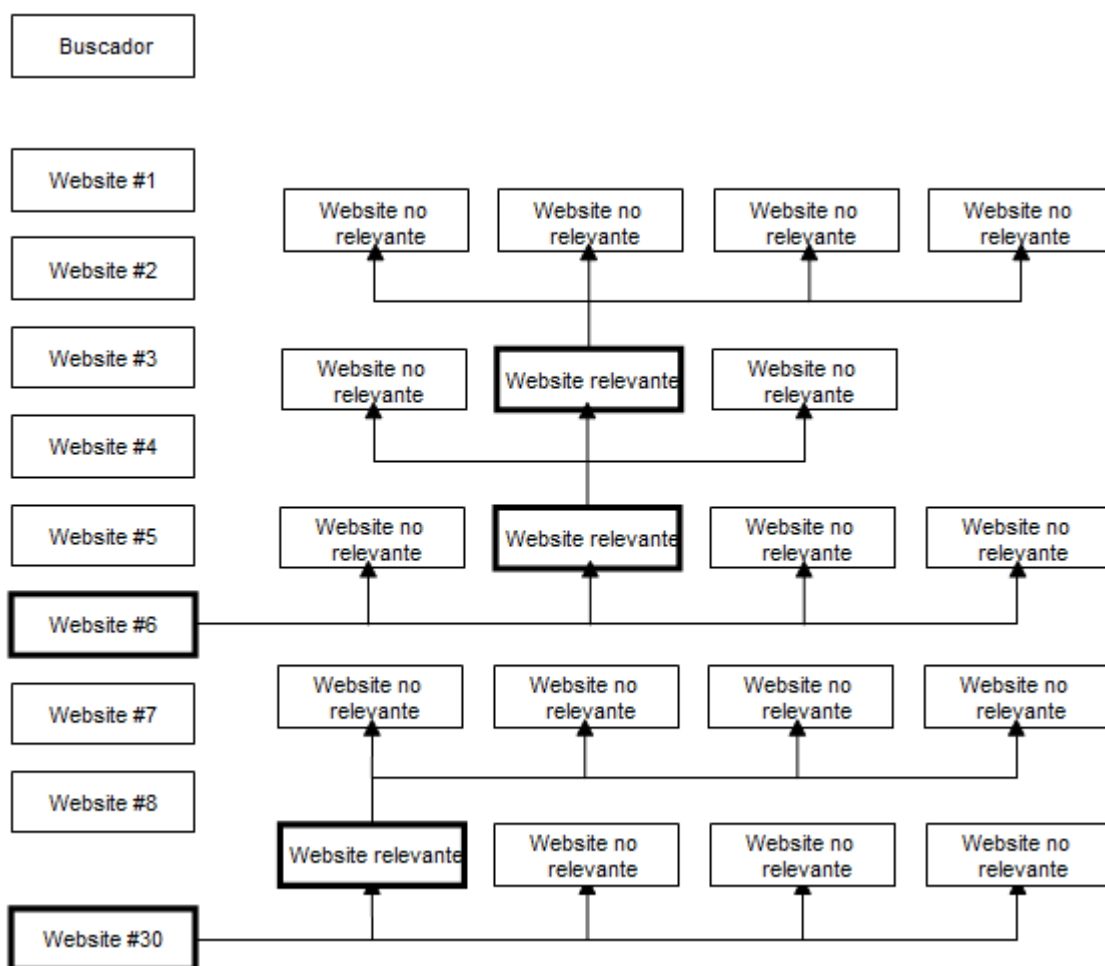
## Appendix A

Término	Definición
Método de Búsqueda	Perspectiva previa al planteamiento de una búsqueda de información concreta que determina, a través de la elección de un método de búsqueda concreto, la forma en la que ésta será llevada a cabo, así como el tipo de resultados que se espera obtener.
Estrategia de Búsqueda	Procedimiento detallado que permite estructurar la secuencia de pasos a seguir al utilizar un determinado método de búsqueda de información. Una estrategia bien estructurada permite asegurar la replicabilidad de la búsqueda, así como acordar razonablemente que las variaciones en los resultados de la búsqueda se deben a una variación en el universo de búsqueda y no en el propio procedimiento.
Instrumento de Búsqueda	Herramienta diseñada para organizar el proceso de búsqueda de información que, en el caso de la búsqueda de información vía web, consta habitualmente de un sistema de entrada, una base de datos sobre la que realizar la búsqueda y un sistema de actualización de su contenido. El tipo de información contenida en su base de datos junto al propio sistema de actualización, determinan la potencia del instrumento y por lo tanto el número de referencias relevantes arrojadas ante una búsqueda determinada.
Ecuación de búsqueda	Término o conjunto de términos considerados relevantes para la localización de fuentes pertinentes de acuerdo a una búsqueda de información concreta, obtenidos tras la adaptación del objetivo inicial de búsqueda al sistema de

	entrada de los propios instrumentos de búsqueda.
Universo de búsqueda	Conjunto global de documentos sobre el que se realizará la búsqueda. Dado que el sistema de incorporación de nuevos recursos no está supervisado por el hombre, la cantidad total de documentos sobre la que se realiza la búsqueda es imposible de conocer, pero queda delimitada teóricamente como la totalidad de recursos a los que existe acceso real vía web. Derivado de esta limitación en su mensurabilidad, los instrumentos de búsqueda trabajan a partir de una base de datos que contiene una representación más o menos ajustada de este universo teórico.

**Tabla 1: Definición de la terminología asociada a la Búsqueda de Información en Internet**

## Appendix B



**Cuadro 1: Esquema de propuesta de estrategia de búsqueda sistemática para la búsqueda exhaustiva de información.**

## Abstract

**Background:** With the progressive increase of publicly available contents on the Internet and as information searching has become one of the most frequent online activities, informal and common searching cannot respond to the aims of scientific research. In regard to the relevant contents in a specific field of knowledge in a non-organized environment such as the Internet, the need to develop efficient procedures to guarantee a systematic search process arises.

**Aim:** The aim of this article is to suggest a framework for developing systematic strategies for the



exhaustive information seeking process on the Internet, easily adaptable to specific seeking goals.

**Method:** Review and proposal of a framework for this decision making process based upon method and instruments selection, and equation development. Furthermore, the whole process is exemplified by the development of a systematic strategy in the field of the information about Health Psychology.

**Conclusions:** The development of these procedures or strategies, articulated by means of the properties of exhaustiveness in the search process, replicability of the procedure and ecologic validity, is not arbitrary, but is directly related to the time spent in the carrying out of the process and the quality of the resulting information.

---

---

---

---