

# Indización y recuperación de la información en Internet. El usuario ante la interpretación y traducción de la demanda

MARTA FUENTES AGUSTÍ  
*Universitat Autònoma de Barcelona*  
JOSÉ HERNÁNDEZ SERRANO  
*Universidad de Salamanca, España*

**L**os ciudadanos de hoy deberán hacer frente a un mundo donde dominar un conjunto de habilidades informacionales básicas resulta imprescindible, al mismo tiempo que deberán convivir en un entorno social y cultural basado en nuevas formas de participación, creación y comunicación caracterizadas por la rápida y constante expansión de las tecnologías de la información y la comunicación (TIC). Esta situación amplía notablemente el campo de investigación en torno a los sistemas de estructuración, indización, recuperación, gestión y organización de la información, al mismo tiempo que convierte la Web en un poderoso instrumento de acceso a la información.

Las autoras (Fuentes, 2006; Hernández, 2006) han observado y analizado cómo principiantes y expertos buscan la información en Internet y la relevancia de ser estratégico en el proceso de búsqueda y recuperación de información. En este proceso una de las fases principales es la interrogación al sistema de información a través de la utilización de la terminología adecuada. La eficacia de esta fase viene determinada por la compleja combinación resultante entre lo que el *usuario* desea encontrar, cómo lo traduce en forma de *pregunta* al

medio, y cómo éste *interpreta* la demanda propuesta por el usuario. Estas tres variables son las que derivan en una respuesta ajustada a la necesidad informativa del momento.

El presente artículo pretende abordar en primer lugar, la comprensión sobre el tipo de información que se halla en Internet; su organización y posibilidades de acceso; para profundizar después en los sistemas prototípicos de recuperación de la información utilizados por los estudiantes de educación secundaria y superior, de acuerdo a los estudios realizados. Por último, concluiremos con algunas consideraciones sobre la formación en procesos alternativos de búsqueda de información, en base a los objetivos de búsqueda, los conocimientos del usuario y la indización del sistema; el uso de términos adecuados y los criterios para su validación; así como la representación e interpretación de la demanda por parte del usuario y del sistema y su incidencia en los resultados obtenidos.

## LA INFORMACIÓN DE INTERNET. INDIZACIÓN Y RECUPERACIÓN.

Uno de los formatos que más se usa en la actualidad para transmitir información es el formato online, basado en las posibilidades que ofrece Internet. En la Red podemos encontrar información de temáticas y índole muy diversa. Por ejemplo, podemos fijarnos en su origen y distinguir diferentes procedencias (institucional, de comunidades locales concretas, gubernamental, de catálogos de bibliotecas, textos impresos digitalizados, de fuentes empresariales y comerciales, de los medios de comunicación, de ciudadanos individuales, etc); en su finalidad y hablar de información comercial, publicitaria, educativa, de ocio, comunicativa, administrativa, de autoprotección, autopromoción, etc.; en su creación y distinguir entre si se desarrolló directamente para la Web, si fue adaptada de una fuente primaria, si es una transcripción literal de una información impresa, etc.

Si nos centramos en la temporalización o vigencia de la información, ésta puede ser histórica, reciente, antigua o actualizada. Por otra parte, puede presentarse en diversos formatos de edición (.doc, .ppt, .pdf, .exe, .html, .zip, .jpg, .gif, etc.); de presentación (textual, hiper-

textual, numérica, visual, sonora, audiovisual, multimedia, etc.) o siguiendo diferentes esquemas, según la forma que ésta puede adoptar (monográficos, ensayos, informes, artículos, documentos académicos, libros electrónicos, comentarios, ponencias, juegos, ficción, sumarios, guías...)

Incluso podemos distinguir la información según su facilidad de localización (organizada y clasificada manualmente, accesible a través de foros, etc.), accesibilidad (de acceso libre o restringido), coste de acceso (gratuita o de pago), etc.

Esta diversidad de contenidos y formatos informativos viene propiciada fundamentalmente por la libertad y simplicidad de publicación y edición que permite Internet. Cualquier usuario de la Red, sin límites ni condiciones aparentes, puede convertirse en autor de la información generando una “realidad” internáutica que puede o no coincidir con la realidad cotidiana, física y tangible, de la información que estamos acostumbrados a transmitir y consumir.

Por un lado, la libertad de publicación propicia un aumento constante de la información a consultar y un amplio abanico de autores procedentes de diferentes organismos o instituciones, geográficamente dispersos con características heterogéneas en cuanto a edad, género o nivel sociocultural. Y por otro lado, la facilidad de publicación que se permite en Internet propicia una rápida edición y reduce la cadena informativa entre la fuente y los usuarios facilitando, a su vez, la actualización instantánea de la información.

Esta situación deja atrás la seguridad de fidelidad y fiabilidad de la información que garantizaban otros medios de información anteriores, permitiendo la circulación simultánea de diversas versiones de un mismo documento, y la desaparición de jerarquías situando en un mismo nivel las producciones de un escritor premio novel con las elaboradas por un estudiante.

Todo esto ofrece mayor visibilidad de la información, desde el punto de vista de su creación, y promueve una nueva identidad y visión de los contenidos, personas, grupos, textos, situaciones, etc. Al mismo tiempo, que permite una mayor circulación de información académica y científica, llegándose a visualizar los *colegios invisibles* de científicos.

Internet es el espacio de difusión de la información más abierto que existe en la actualidad, de fácil acceso, a-temporal y a-espacial, transversal, fácil de usar, atractivo, etc. pero, al mismo tiempo, se caracteriza por su *desorden documental*. Nació en el Laboratorio Europeo de Física de Partículas del Centro Europeo de Investigación Nuclear (CERN) con una finalidad muy distinta a la de transmitir información a escala mundial (Vea, 2005). Esto ha propiciado que no exista un control de organización de la información.

El incremento y falta de organización de la información, su diversificación y su localización distribuida, genera una serie de circunstancias y consecuencias que deben ser valoradas. Entre ellas tomamos en consideración la relación entre la facilidad de publicación y una posible sobreinformación —que dificulta la selección— e infoxicación —que dificulta la valoración de la credibilidad y fiabilidad de la información— junto con la inminente caducidad de misma (mucha información esta sin actualizar o no se sabe cuándo se actualizó por última vez).

Ante esta estructura informacional resulta imprescindible el uso de herramientas para poder localizar la información precisa en el momento oportuno. La Red nos ofrece diferentes posibilidades de acceso a la información que van de las más tradicionales (consulta de una base de datos recopilados y clasificados manualmente) a los más sofisticados motores de búsqueda nacidos en la era de Internet.

## HABILIDADES INFORMACIONALES PARA RECUPERAR LA INFORMACIÓN PUBLICADA EN LA RED

Venimos hablando de un formato para transmitir la información remoto, dinámico y en constante evolución que requerirá unos usuarios con habilidades informacionales generales y específicas (Fuentes, 2001).

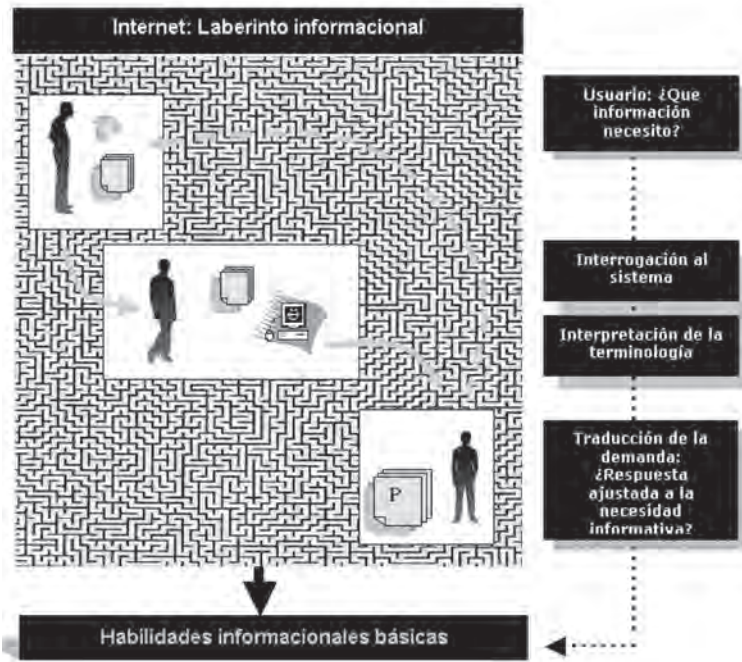
El creciente intercambio de información económica, jurídica, deportiva, científica, artística, cultural, etc. sin una indización aparente, ha creado la necesidad de contar con sistemas de recuperación de información cada vez más sofisticados y diversos de índole muy diversa. Podemos hablar de tres posibilidades para la recuperación de la información publicada en la Red.

1. Mediante búsquedas realizadas en bases de datos tradicionales perfeccionadas con el avance de las Tecnologías de la Información y la Comunicación (TIC): bases de datos de las bibliotecas con acceso remoto, bases de datos referenciales, hemerotecas digitales, bases de datos a texto completo, especializadas, integrales, etc.
2. Mediante herramientas que permiten búsquedas directas en la red de Internet generalmente a través de robots: directorios o índices temáticos, motores de búsqueda genéricos o específicos, metabuscadores, portales.
3. Mediante herramientas y sistemas de comunicación interpersonal y trabajo colaborativo que nos ofrece la red: correo electrónico, chat, videoconferencias, foros, debates, listas de distribución, etc.

Las características de estos medios de búsqueda incrementan la necesidad de contar con estructuras terminológicas precisas para recuperar adecuadamente los contenidos deseados. No basta con seguir un proceso automatizado de búsqueda caracterizado por un *zapping* compulsivo sin partir de un criterio previo de selección sino que se precisa de un proceso deliberativo (Monereo y Fuentes, 2005).

Los expertos en la temática identifican varias fases en el proceso de recuperación de información en Internet (Fuentes, 2006, 309-327). A pesar de las peculiaridades de cada uno de sus modelos coinciden en resaltar la importancia de la definición de la demanda, la elección del medio y lenguaje de búsqueda, la selección de la información hallada y la interpretación y utilización de esta información. Centrándonos en el proceso de recuperación de la información debemos insistir en el laberinto informacional que supone Internet en cuanto a la falta de indización de la información y la exigencia de ciertas habilidades informacionales básicas. Habilidades que deberían poseer todos los usuarios de la red.

Figura 1. Recuperación de la información en Internet. De la necesidad informativa a los contenidos.

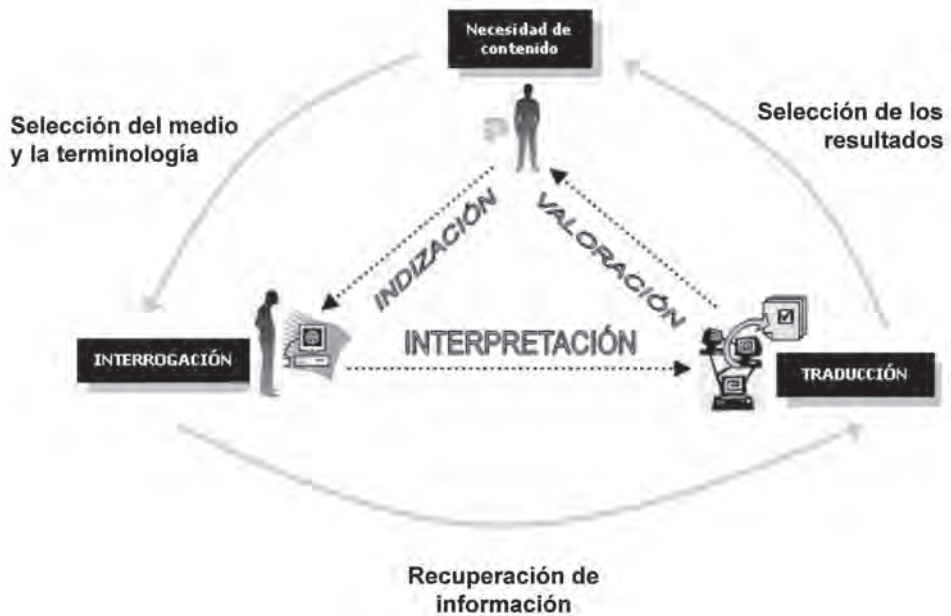


En la figura precedente hemos sintetizado el proceso de recuperación de la información haciendo hincapié en la necesaria *interpretación y traducción de la demanda por parte del usuario* el cual recibirá como respuesta la *interpretación y traducción del sistema* en un listado de posibles recursos, fuentes informativas, etc. que pueden o no dar respuesta a su necesidad.

Para que el sistema proporcione la respuesta deseada se debe interrogar correctamente al sistema de búsqueda elegido. Por un lado, se debe tener presente las características y modo de búsqueda del sistema y por otro las características de la información que se desea hallar. Con todo ello, se debe definir muy bien la demanda usando la terminología adecuada, ampliando la elección de las palabras clave a la elección de los conectores entre ellas y/o las opciones limitadoras de la búsqueda que nos ofrece el sistema.

Podemos hablar de tres grandes variables que inciden directamente en esta parte del proceso de búsqueda de información en la red de Internet: lo que el usuario desea encontrar, la pregunta que se formula al sistema y la interpretación que hace el sistema a dicha pregunta.

Figura 2: Contenido, pregunta y respuesta en la recuperación de información en Internet.



Cuando el usuario decide realizar la búsqueda en la red de Internet debe plantearse que tipo de recurso necesita (un documento de texto, una imagen, una hoja de cálculo, un video, etc. divulgativo, académico, etc. actual, etc.), cual será el sistema de búsqueda más apropiado (base de datos estructurada, motor de búsqueda genérico o especializado, directorio, metabuscador, lista de distribución, consulta a un experto mediante el correo electrónico, portal, etc.) y de entre estos el más adecuado (Google, Altavista, Yahoo, la web del ministerio, de

una escuela, etc.) para llegar al contenido deseado.

Esta premisa nos servirá de punto de partida para la exploración, pero una vez elegido el sistema de búsqueda (base de datos estándar, motor de búsqueda, directorio, etc.) la eficacia viene determinada por la compleja combinación resultante entre lo que el usuario desea encontrar, cómo lo traduce en forma de pregunta al sistema elegido, es decir, la traducción de su demanda al lenguaje del sistema para que este lo interprete adecuadamente y lo traduzca en un listado limitado de posibles documentos o enlaces que satisfagan la necesidad informativa del usuario (figura 2).

Si se usa el lenguaje de interrogación adecuado el sistema interpretará correctamente la demanda y este recuperará la información de manera eficaz. En caso contrario se deberá desestimar el resultado y repetir la búsqueda. Aconsejando no iniciarla de nuevo, sin más, sino actuar de manera estratégica (consciente e intencionalmente) y autorregular el proceso.

Los enlaces recuperados por el sistema deben de ser un listado limitado. Recuperar información en Internet no consiste en localizar información sobre la temática deseada sino que esta pueda dar respuesta a nuestros objetivos. Es decir, cumpla con los requisitos mínimos para que el usuario pueda asumir la selección del o los recursos idóneos teniendo en cuenta el número de resultados, el ajuste al tópico de la búsqueda, la credibilidad y fiabilidad de la información hallada, etc.

Elegir el camino correcto no es una tarea fácil y requiere de ciertos conocimientos y habilidades informaciones que desde la educación obligatoria u otras instituciones (por ej. las bibliotecas) se debe tratar de ayudar a adquirir mediante tareas auténticas.

Nuestros estudios (Fuentes, 2006; Hernández, 2006) han identificado entre los estudiantes seis modalidades de búsqueda de información. De estas modalidades las más usadas se alejan bastante de las usadas por los expertos. Este hecho se debe principalmente al tipo de tareas planteadas habitualmente en la escuela obligatoria o en la educación superior.

Generalmente los estudiantes han recuperado información de Internet para dar respuesta a demandas simples con lo que un proce-



so automatizado puede llegar a satisfacer su necesidad informativa. Mientras que ante una demanda más compleja se ha evidenciado la necesidad de ser estratégico en la búsqueda y las modalidades más habituales de búsqueda les han resultado infructuosas.

Con todo queremos insistir una vez más en la importancia de la interpretación y traducción de la demanda, en la necesidad de ser consciente de la relevancia de las decisiones tomadas (elección del sistema de búsqueda y de la terminología usada) y de todo el proceso que implica recuperar información de la red. Información que debe de ser seleccionada en base a unos criterios para pasar a ser analizada, tratada y usada adecuadamente.

## A MODO DE CONCLUSIÓN

Conocedores de la organización del conocimiento, los procesos de recuperación de la información y la rápida y constante evolución de los formatos de transmisión se debe proceder, por un lado, a analizar como acercar a los ciudadanos las habilidades informacionales básicas argumentadas y, paralelamente, analizar los métodos utilizados para elaborar herramientas de indización y recuperación de la información, las técnicas utilizadas para el tratamiento terminológico, la representación del contenido temático de los documentos multimedia, las normas y opciones para organizar las estructuras temáticas y terminológicas en diferentes ámbitos del conocimiento, los programas para la gestión y clasificación de la información recuperada, los problemas no resueltos de acceso a formatos diversos de transmisión de la información (brecha digital), de compatibilidad entre los sistemas de almacenamiento, indización y recuperación de la información, etc.

Sin pretender ser más exhaustos en la temática, concluimos este artículo insistiendo en que ante un banco creciente y sin control aparente de información se requiere, más que en cualquier otro formato de recuperación de información, adquirir ciertas habilidades informacionales generales y específicas para poder ser estratégico en la búsqueda y selección de información.

## REFERENCIAS BIBLIOGRÁFICAS

- Fuentes, M. (2001). Naufragar en Internet. Estrategias de búsqueda de información en redes telemáticas. *Virtual Educa*, Madrid, 27-29 de junio de 2001. [Publicado por la UOC en <http://www.uoc.edu/web/esp/art/uoc/0109037/fuentes.html#bibliografia>]
- Fuentes, M. (2006). *Estratègies de cerca i selecció d'informació a Internet. Anàlisi de les modalitats de cerca i selecció d'informació a Internet dels estudiants de quart curs d'educació secundària obligatoria*. UAB: Tesis Doctoral.
- Hernández, M. J. (2006). *Estrategias de búsqueda de información en Internet en estudiantes universitarios*. USAL: Proyecto de investigación doctoral.
- Monereo, C. y Fuentes, M. (2005). Aprender a buscar y seleccionar en Internet. En Monereo, C. (Coord.). (2005). *Internet y competencias básicas*. Barcelona: Graó. 27-50.
- Vea, A. (2005). *Who is Who in the Internet World?*. Barcelona: CIDEM.