

Aunque hay más de 80 millones de internautas hispanohablantes, que representan en torno al 8% de todos los que en el mundo utilizan internet con mayor o menor frecuencia, sólo es posible encontrar indexados en español un 2 por ciento de todos los contenidos digitalizados por los buscadores en la Red, mientras que los contenidos en inglés superan el 68%. Esta desproporción afecta negativamente, entre otros, al desarrollo de nuestras industrias cultural y turística. Para prevenir y solventar esta problemática ha nacido el proyecto «Existes», un proyecto de buscador federado español, cuyas características describe la siguiente nota (1)

«EXISTES», UN PROYECTO DE BUSCADOR FEDERADO PARA CONTENIDOS DE INTERNET EN ESPAÑOL

La cultura española se encuentra actualmente con dos grandes problemas en relación con el vasto mundo de Internet: uno, relativo al idioma y otro concerniente a los iconos. Estos problemas afectan negativamente a la promoción en el mundo de nuestra cultura y nuestro patrimonio cultural y al desarrollo del turismo en nuestro propio país.

EL IDIOMA

El primero de los problemas es la escasa presencia relativa del idioma español en los grandes motores de búsqueda (como Google, Yahoo, etc.), en relación con su peso real en número de internautas y volumen de contenidos en la Red.

Actualmente hay más de 80 millones de internautas hispanohablantes en el mundo, es decir, hispanohablantes que navegan por la Red con cierta frecuencia, y que representan en torno al 8 por ciento de todos los internautas del mundo (2). Sin embargo, los contenidos en español que están indexados por los grandes motores de búsqueda sólo representan el 2 por ciento del volumen total de información indexada

por esos buscadores, mientras que los contenidos en inglés representarían en torno al 68 por ciento del total (3). Ello quiere decir que, cuando un internauta anglófono utiliza un motor de búsqueda como Google, dispone de muchísimos más documentos para los que obtiene respuesta a sus preguntas o búsquedas que un hispanohablante que hace la misma búsqueda o pregunta en español.

Al respecto, cabría pensar que los anglófonos escriben más que los hispanohablantes y que, por tanto, tienen más contenidos en la Red. Pero lo cierto es que la producción de contenidos en español es proporcional al número de internautas hispanohablantes; y sin embargo, la presencia de nuestro idioma en los motores de búsqueda no se da, ni mucho menos, en la misma proporción.

Esa escasa presencia del español en los grandes motores de búsqueda puede explicarse básicamente por tres razones. En primer lugar, porque se trata de empresas privadas que, como tales, han de tener rentabilidad y beneficios, y éstos provienen esencialmente de los ingresos por publicidad. Google, por

ejemplo, tiene mucho más mercado en los países anglosajones, donde la cultura de la inversión publicitaria *online* se ha desarrollado más; y tiene muchos menos intereses en los países latinos, donde la cultura de la publicidad *online* se ha desarrollado menos (4).

La segunda causa de la escasa presencia del español en los grandes motores de búsqueda reside en las economías de escala: Indizar contenidos en inglés tiene una enorme ventaja, ya que es la segunda lengua de muchos millones de personas en el mundo. En consecuencia, al indizar contenidos en inglés se está abarcando un espectro mucho más amplio de usuarios potenciales, lo que no sucedería si la indexación se centrara exclusivamente en contenidos en castellano. En este caso se dejaría fuera, ajenos a esa indexación, a un volumen muchísimo mayor de internautas no hispanohablantes y cuyo segundo idioma es el inglés.

La tercera razón es que la capacidad de los buscadores es limitada: no pueden indizarlo absolutamente todo. Si los grandes motores de búsqueda tuvieran capacidad para indexar todos los contenidos publicados o vertidos a la Red, no habría problema: indizarían todo y facilitarían a cada internauta o usuario lo que pidiese o buscara, y en el idioma disponible. Pero, hoy por hoy, la capacidad de crecimiento (indexación) de los motores de búsqueda es mucho más limitada que la velocidad a la que crece la Red (contenido). Hay quien asegura que la Red duplica su tamaño (contenidos) aproximadamente cada ocho meses! Y, obviamente, un buscador no puede absorber toda esa nueva información en ese lapso de tiempo, no puede crecer a esa velocidad.

Este es, pues, el escenario en el que se plantean los problemas del idioma español en la Red (5). Precisamente en el momento mismo en que se está tratando de potenciar la Sociedad de la Información, los usuarios españoles (y, en definitiva, los hispanoparlantes) estamos menos motivados que los anglófonos para utilizar la Red, porque tenemos muchos menos contenidos a nuestra disposición.

LOS ICONOS

El segundo problema de la cultura española en Internet concierne a los iconos culturales, esto es, aquellas palabras clave que identifican contenidos de referencia de la cultura española.

Picasso, por ejemplo, es uno de los principales iconos culturales españoles en la Red, si es que no el español actualmente más buscado en Internet de entre todos los referentes pictóricos del siglo XX. Por ello,

nuestro país debería ser capaz de recibir prácticamente toda la demanda de contenidos que se generara sobre Picasso; y también debería ser el principal oferente de información sobre el mismo. Esto podría generar beneficios en diversos terrenos, como en el turismo cultural e idiomático (6), y proporcionarnos capacidad de influir en ámbitos tales como la enseñanza del arte y la cultura (7).

Pero la realidad es que si buscamos Picasso en Google nos encontramos con que hay, por ejemplo, 15 millones de documentos en inglés con contenido sobre Picasso, mientras que sólo hay 2,6 millones de documentos en español (8). Y si se busca Picasso en Nueva York, el buscador inmediatamente da la respuesta en inglés (aunque se esté buscando en español), porque es mucho más potente la información que hay en inglés que la que hay en castellano.

En principio, si esa información en inglés la hubiera facilitado una entidad cultural española no debiera preocuparnos gran cosa. Pero lo normal, por el contrario, es que esa información en inglés sobre Picasso haya sido producida en EE.UU y esté controlada por entidades culturales o empresariales estadounidenses, las cuales facilitarán la información que desean se consuma sobre Picasso, y no necesariamente la información que desde España se hubiera deseado trasladar al público sobre el artista, su obra u otras cuestiones relacionadas.

Las consecuencias de lo anterior pueden ser de diverso tipo y alcance. A título de ejemplo, esa situación puede tener importantes repercusiones en el turismo. En efecto, si un ciudadano de California busca Picasso en Internet, es muy posible que el buscador le remita antes a un cuadro del artista malagueño en una exposición de Nueva York, que a una exposición retrospectiva de la obra de Picasso en Málaga (9). En consecuencia, el ciudadano californiano viajaría, en su caso, a Nueva York para visitar la exposición neoyorkina, en lugar de desplazarse a la exposición retrospectiva de Málaga y hacer turismo en la Costa del Sol..., con las consiguientes implicaciones.

LOS BUSCADORES LOCALES

El bajo volumen de indexación en español por parte de los grandes buscadores y, en consecuencia, la escasa presencia de nuestro idioma en Internet son problemas en los que difícilmente podemos influir a corto plazo. En cambio, sí podemos realizar grandes avances ideando soluciones adecuadas a los problemas relacionados con los buscadores locales de los portales (*sites*) españoles pertenecientes, en nuestro caso, a instituciones y entidades de la esfera pública, como son los portales de las Administraciones

Públicas, las entidades de promoción de la cultura y el turismo, y las Universidades.

En general, los portales de esas entidades y organismos disponen de un motor de búsqueda propio en cada portal (buscador local). Pero esos motores de búsqueda son heterogéneos entre sí en cuanto a calidades, capacidades y comportamientos; con bastante frecuencia no recuperan o no devuelven muchos de los documentos existentes en el propio repositorio del portal, ni ofrecen alternativa alguna cuando el documento o la información deseada no se halla dentro del portal en cuestión; tampoco disponen de las funcionalidades más avanzadas ni de una herramienta de retroalimentación (*feedback*) que facilite el conocimiento de los gustos, inquietudes e intereses de los internautas; además, y más grave todavía, cada portal funciona y permanece como un compartimento estanco y aislado del resto de portales en lo relativo a su fondo documental propio o específico y a las directrices relativas a la indización del mismo.

De manera que el funcionamiento aislado e independiente de cada uno de estos buscadores locales (de exclusiva utilidad para el portal de que se trate) supone desperdiciar un valiosísimo capital colectivo en perjuicio de los propios usuarios y de la sociedad en general. El buscador local sólo da cuenta de la información disponible en el propio portal, y ello en el supuesto de que aquella esté convenientemente indizada y sea posible su recuperación; deja búsquedas sin respuesta o no proporciona alternativas cuando la información buscada no se encuentra en el portal en cuestión (no contesta a la pregunta esencial que los internautas hacen a un buscador: «¿Dónde se encuentra lo que yo necesito?»); no remite a otro portal donde la información buscada pueda estar disponible, ni siquiera cuando se trate de una institución «hermana» —como puede ser otra Universidad u otro Ministerio—, entre otras disfunciones y deficiencias.

«EXISTES», UN BUSCADOR FEDERADO

Estas cuestiones han venido siendo objeto de consideración, desde hace varios años, por un equipo informal de expertos en Tecnologías de la Información (10). Dichos expertos han llegado a la conclusión de que las mencionadas deficiencias pueden corregirse de manera sustancial haciendo que el mecanismo de búsqueda de todos y cada uno de los portales pertinentes (Administraciones Públicas en sus diferentes niveles, entidades de promoción cultural y turística, Universidades) opere de forma conjunta (mediante la agregación *inteligente* de los buscadores de cada portal, para configurar de esta manera un buscador federado).

Consecuentemente, cualquier búsqueda se efectuaría sobre la base documental de todos y cada uno de dichos portales, lográndose así un sistema de conocimiento común y la interconexión de los resultados. Se aprovecharía de esta manera un capital enormemente valioso —hasta ahora totalmente desaprovechado— y se ofrecería a los usuarios un resultado mucho más efectivo y funcional.

Es así, sobre la base de esas reflexiones y en esa perspectiva, como ha nacido el proyecto de un motor de búsqueda federado y distribuido entre los portales pertinentes, que ha recibido la denominación de Proyecto «*Existes*». Como motor de búsqueda, cumple la función de rastrear, analizar e indizar, primero, para poder dar respuesta después a las consultas de los usuarios.

Es distribuido porque se integra en los portales de cada una de las entidades que participan en el proyecto (está repartido o distribuido entre todos los portales mencionados; de hecho, es un agrupamiento *inteligente* de los buscadores locales de todos los portales participantes). Y permite federar los resultados porque cuando un usuario plantea una pregunta al buscador dentro de uno de esos portales, su pregunta es replicada para cada uno de los portales de la federación, y la respuesta se construye a partir de la suma de todos los recursos, por lo que puede localizar los contenidos de toda la red de portales que comparten este servicio.

La estructura federada contaría con tantos portales como entidades o instituciones quisieran sumarse al proyecto de motor de búsqueda, pudiendo accederse al buscador federado desde cualquier portal de la red de entidades federadas; y, por consiguiente, desde cualquiera de estos portales sería posible acceder, a través del buscador, a la totalidad de los contenidos de la federación. Dicho de otro modo, el modelo federado inicia la búsqueda de información dentro de cada portal (buscador local); y a partir de cada uno de los «buscadores locales» y agregándolos de modo *inteligente* (federándolos) proporciona un «buscador global» que suministra resultados de todos los portales o *sites* federados. Ante las búsquedas de los usuarios, ofrece resultados procedentes del propio portal (búsqueda local) y resultados de la red federada (búsqueda global).

Un motor de búsqueda federado es una solución inteligente y colaborativa, de bajo coste, con ventajas para todas las partes implicadas, no precisa desplazar a ninguno de los grandes buscadores ya existentes, garantiza la indización de la totalidad de los contenidos del portal federado, permitiendo así aflorar muchos contenidos que hoy permanecen ocultos;

ofrece un mejor servicio al internauta, ya que éste puede acceder al conjunto de los recursos de la federación a través de cualquiera de los portales federados, y permite disponer de una potente herramienta de conocimiento para analizar las necesidades y orientaciones de los usuarios, herramienta utilísima de la que hoy se carece.

Para el internauta, su funcionamiento es muy sencillo: El usuario no tiene más que entrar directamente en el portal federado que le interese y utilizar el buscador de ese portal para hacer su búsqueda. De esta manera, se encontraría directamente dentro de una red federada —formada por el conjunto de todos los buscadores locales de los diferentes *sitios* web de las entidades integrantes de la federación— que atenderían su búsqueda o consulta, que podría ser tanto de alcance local, (concerniente sólo al contenido del portal concreto al que el usuario ha accedido) como global, (relativo al contenido de cualquiera de los portales federados) (11).

En principio, esta solución podría no tener un impacto importante sobre el tráfico (no modificaría sensiblemente el actual volumen de tráfico en la Red), pero en cambio contribuiría sensiblemente a la maduración o mejora de la estructura de los portales (*sites*) españoles de la esfera pública. En efecto, la creación de una estructura federada de motores de búsqueda hace posible que cada *webmaster* pueda controlar los rastreos, las tareas de indización que se realizan en su portal, conocer los problemas (barreras) que presenta su portal en relación con la indización los motores de búsqueda, madurar su posicionamiento en la Red y acometer otras mejoras.

El proyecto de buscador federado ha tenido una amplia aceptación, además de por las razones ya señaladas, porque ninguna de las instituciones o entidades participantes tiene un control dominante, sino que, no obstante la puesta en común de los recursos federados, cada entidad colaboradora mantiene el control de sus propios servicios y su propia autonomía.

LA EXPERIENCIA PILOTO

Para llevar a la práctica la idea del buscador español federado se ha contado con el apoyo tecnológico del Grupo ALMA (12), al que, inicialmente, se le ha encomendado la realización de un primer proyecto piloto, puesto en marcha en febrero de 2006 (13).

La experiencia piloto tiene tres objetivos principales: experimentar el modelo de explotación del motor de búsqueda federado, conocer la adecuación técnica de la solución propuesta a las necesidades de cada

entidad participante y plantear, al término de la experiencia, un diseño avanzado de cómo debiera ser el servicio en explotación.

La experiencia ya en marcha abarca tres ámbitos diferenciados de la esfera pública: Administración General del Estado, universidades, y cultura y turismo. En cada uno de esos tres ámbitos, las entidades participantes en la experiencia piloto están agrupadas en un mismo nodo, lo que facilita la experimentación independiente de cada nodo y, por consiguiente, también de las entidades adscritas a cada uno de los tres ámbitos señalados. Es decir, se ha optado por empezar con tres ensayos parciales (tres ámbitos muy específicos con sólo algunas entidades participantes en cada uno de ellos) (14). La experiencia piloto cuenta con el compromiso de participación de una treintena de portales, lo que supondría un volumen de indización de 10 millones de páginas y un tráfico de unas 60.000 consultas al día.

Esta prueba piloto permite ya al usuario final utilizar el buscador (local) de cualquiera de las entidades participantes y consultar contenidos no sólo en cada portal individual sino —desde el portal al que se ha accedido— en todo el conjunto de portales federados (búsqueda global). Al mismo tiempo, permite a los *webmasters* de cada entidad participante evaluar el impacto de la prueba piloto en su propia estructura y conocer de cerca las ventajas de participar en esta red de búsqueda federada. Del resultado de esta experiencia dependerá la solución definitiva que se dé al buscador federado.

El primer ámbito de la experiencia piloto en marcha es el que se ha denominado Nodo de Cultura y Turismo, el más avanzado de los tres actualmente en experimentación. En el mismo participan una decena de organizaciones o entidades: Casa de América, Real Instituto Elcano, Red.es, Sociedad Estatal de Gestión de la Información Turística (SEGITUR); Organización de Estados Iberoamericanos para la Educación, la Ciencia y la Cultura (OEI); Museo Thyssen-Bornemisza, RTVE, Secretaría General Iberoamericana (SEGIB), Promoción Madrid y Consejo Superior de Investigaciones Científicas (CSIC).

Los contenidos de esas entidades han sido indizados por un motor de búsqueda central y ya están siendo objeto de explotación por los interesados. Un internauta o usuario que entre en la página web de, por ejemplo, RTVE en busca de determinada información puede encontrar esa información aunque no esté en la página de RTVE sino en la de Casa de América, en el Museo Thyssen, en Red.es o en cualquiera de las restantes entidades participantes, o a la inversa. Así es como funciona el buscador fe-

derado: busca la información solicitada en la propia página web visitada, pero también y adicionalmente en todos los demás portales de la estructura que se ha federado, en toda la red de instituciones y entidades que comparten el motor de búsqueda.

Paralelamente a ese primer nodo, se está poniendo en marcha un segundo nodo integrado por varios Departamentos de la Administración General del Estado: (AGE): en un principio, Ministerio de Administraciones Públicas, Ministerio de Industria, Turismo y Comercio; Ministerio de Defensa y Ministerio de Fomento. Se ha construido este segundo nodo específico y separado del primero precisamente por las especiales características de la Administración en cuanto a tipo de información y documentación, ritmos, procedimientos, dinamismo, grado de agilidad, etc., diferencias éstas que aconsejaban llevar esta experiencia piloto con un nodo diferente.

Pero la idea central, dentro del Nodo AGE es básicamente la misma: que cuando un internauta busque una información en la página web de un Ministerio cualquiera (por ejemplo, el MAP) y éste no disponga del contenido buscado pero sí esté en otro Ministerio (por ejemplo, Defensa), el ciudadano no se quede sin contestación, sino que obtenga la respuesta a su búsqueda ya en la página del primer Ministerio al que accedió.

Esta experiencia piloto se completa con un tercer nodo, todavía en fase embrionaria, el Nodo Universidades, en el que están colaborando, por el momento, las de Educación a Distancia (UNED) y la de Santiago de Compostela. Conviene señalar, en este contexto, que las universidades son las principales generadoras de contenidos en nuestro país. Tan sólo la UNED tiene del orden de los 12 millones de páginas en castellano, esto es, aproximadamente el 5 por ciento de las páginas en español con que cuenta Google (unos 250 millones en el momento de redactar estas líneas). Y la UNED es sólo una de las 976 universidades que están dentro del vasto proyecto *Universia* (15).

El liderazgo del proyecto *Existes* corresponde al Ministerio de Administraciones Públicas (MAP), que es el primer interesado en que funcione bien, en que se pongan en marcha este tipo de herramientas de colaboración que benefician, en última instancia, a todos los ciudadanos. Al Ministerio de Industria, Turismo y Comercio (MITYC), por su parte, le compete el terreno de las infraestructuras correspondientes para que el proyecto sea viable.

Esta experiencia cuenta, además, con dos observadores excepcionales. Por una parte, el Departamento

de Lenguajes y Sistemas Informáticos de la UNED, uno de los más avanzados de España en el análisis de motores de búsqueda; y por otra, el Centro de Información y Documentación Científica (CINDOC), del CSIC, una de las entidades más experimentadas en optimización de buscadores (SEO) (16) y madurez de páginas web. Ambas instituciones están colaborando en la supervisión de la experiencia piloto.

En resumen, está ya en funcionamiento un proyecto piloto de instituciones y entidades de la esfera pública que están probando una iniciativa de motor de búsqueda federado que, en parte, pretende afrontar la problemática específica del idioma español y de la cultura hispana en Internet señalada al principio de esta Nota. Este dispositivo permite a los *webmasters* madurar y mejorar sus posicionamientos en la Red, conocer lo que se está buscando en ella y, por tanto, orientar sus contenidos hacia esa demanda, teniendo en cuenta el enorme potencial que representa la comunidad de internautas hispanohablantes, y en particular la comunidad iberoamericana.

OPORTUNIDADES Y RETOS

Como consecuencia de la experiencia piloto que se está llevando a cabo, los *webmasters* de las entidades participantes en el proyecto están tomando conciencia de las barreras que hasta ahora existían en sus páginas web para que un buscador como Google pudiera indizar sus contenidos, con independencia de que fuera ese el objetivo o las instrucciones del buscador. Esta toma de conciencia está permitiendo a los responsables de los portales asociados superar esas barreras y, consecuentemente, mejorar su posicionamiento en la Red, simplemente por suprimir aquéllas y hacer posible que sus páginas web sean ahora accesibles a la indización de los grandes motores de búsqueda (17).

Precisamente, una gran ventaja del modelo de estructura federada es que se garantiza la visibilidad de todos los contenidos, desapareciendo así el problema de la «web invisible» (18), porque se cuenta con el propio interés de los *webmasters* de cada uno de los portales asociados para que haya absoluta visibilidad de los contenidos. Con ello se garantiza la indización total de los contenidos de la red de estructura federada.

El reto está en conseguir que la estructura federada tenga un crecimiento rápido, al objeto de alcanzar cuanto antes una masa crítica de contenidos, usuarios y entidades participantes (portales), que se sumarán al proyecto ante las perspectivas de más tráfico, mayor conocimiento y nuevas ventajas.

El proyecto ha generado grandes expectativas. En primer lugar, porque podría colocar a España en una destacada posición en Internet, con capacidad para atender a una comunidad hispanohablante de internautas muy superior a la población española propiamente dicha, y situarla entre los países tecnológicamente más avanzados de la comunidad hispanohablante.

En segundo lugar, porque las implicaciones económicas del proyecto son considerables. Basta con hacer referencia al importante volumen de recursos que actualmente se destinan a conseguir que nuestras instituciones tengan una mayor presencia en Internet y a que los ciudadanos entremos en la Sociedad de la Información, al mismo tiempo que los contenidos —que son la materia prima— permanecen en compartimentos estancos, absolutamente ocultos, sumergidos, de hecho como si no existieran.

De manera que podría decirse que se está desperdiciando buena parte de la inversión que se está haciendo en este terreno, porque esos contenidos no afloran, al no haber sido indizados por los grandes motores de búsqueda (19). De ahí la imperiosa necesidad de llevar a buen puerto este proyecto de motor de búsqueda federado.

España es actualmente un país consumidor de contenidos, cuando debiera ser un generador neto de los mismos, tanto por la importancia del idioma español (20) como por la existencia de una vasta comunidad hispanohablante. En cambio, Francia y Alemania, con un potencial mucho menor de internautas en sus respectivas lenguas, son generadores de contenidos y han sentido la necesidad de crear su propio buscador, el proyecto «Quaero» (21). Cabe pensar, entonces, cuánta mayor necesidad tendría España de hacer lo propio para los contenidos en castellano. Y tanto más aún por cuanto que Francia y Alemania parten de una mejor posición que España, ya que son generadores netos de contenido mientras que nosotros somos consumidores.

Convendría aprovechar esta oportunidad y evitar así que, dentro de unos pocos años, la capacidad potencial de España para manejar la cultura hispana en la Red quedara definitivamente en manos de otros países. En tal caso, España se limitaría a consumir contenidos generados en esos países, manteniendo sólo una muy modesta capacidad propia de generación. Por todas estas razones, el proyecto constituye un importante reto y una oportunidad merecedora del mayor apoyo.

LUIS PORTILLO

NOTAS

- (1) El autor agradece la inestimable ayuda prestada por Chimo Soler (Instituto Cervantes) y Adolfo Corujo («ALMA Technologies, S.A.»), absolutamente imprescindible para la redacción de esta Nota.
- (2) <http://www.internetworldstats.com/stats7.htm>
- (3) <http://global-reach.biz/globstats/refs.php3>
- (4) Así, en EE.UU —que es el mayor inversor en publicidad per cápita del mundo—, aproximadamente el 7 por ciento de toda la inversión publicitaria del país va a la publicidad *online* y, dentro de ella, el 50 por ciento va a motores de búsqueda. En consecuencia, Google tiene unos enormes ingresos por publicidad en EE.UU. En España, en cambio, sólo se dedica a publicidad *online* en torno al 1,5 por ciento del gasto publicitario total; y de ello, sólo el 20 por ciento a publicidad con motores de búsqueda.
- (5) Algunos expertos aducen una razón más de la escasa presencia del castellano en la Red: el hecho de que nuestras páginas web, nuestros portales (*websites*) sean poco maduros desde el punto de vista de la orientación hacia motores de búsqueda; esto es, presentan deficiencias y barreras para que sus contenidos puedan ser objeto de indización por los motores de búsqueda. Es decir, se trataría no ya sólo del problema de que los grandes buscadores, como Google, indexen o no en español, sino también, y además, de un problema de madurez de nuestra propia estructura.
- (6) Según cifras estadísticas, el 60% de las decisiones de turismo cultural se toman por Internet. *Iconos Culturales Hispanos en Internet*, en: <http://www.alasybalas.com/archives/000113.html>
- (7) *Arte y cultura para nuevos públicos en la era digital*, <http://www.cervantes.es/uimp2004/>
- (8) España ha dejado en manos de los demás patrimonio cultural propio. Esto ha supuesto hechos como que Dalí tenga aproximadamente tres veces más documentos en inglés que en español. Otro ejemplo es que Goya sea más referenciado en museos extranjeros que en el propio Museo del Prado. *Iconos Culturales Hispanos en Internet*, en: <http://www.alasybalas.com/archives/000113.html>
- (9) Esta cuestión remite al tema del «posicionamiento en Internet» (el lugar en que se sitúa una determinada página web en el ranking del buscador). http://es.wikipedia.org/wiki/Posicionamiento_web
- (10) Directores de Informática e Internet de entidades tales como Instituto Cervantes, CSIC, Casa de América, Red.es, RTVE, UNED, SEGIB, Real Instituto Elcano o Universidad de Santiago de Compostela, entre otras.
- (11) Si el usuario entrara, por ejemplo, en la web del Instituto Cervantes y preguntara al buscador de esa página, obtendría contestación no sólo respecto a los contenidos de la web del Instituto Cervantes, sino sobre la base de todos los contenidos de todas las páginas web indizadas por el motor de búsqueda federado, esto es, la base documental de todo el conjunto de portales (*websites*) asociados a la iniciativa federada (Casa de América, RTVE, Ministerio de Fomento, etc.).
- (12) <http://www.almatech.es/> - <http://www.grupoalma.com/>
- (13) En realidad, se trata de un proyecto de buscador iberoamericano federado, dirigido a los 96 millones de internautas que buscan contenidos en castellano, portugués, catalán, gallego y vasco. <http://www.existes.net/existes/>
- (14) En estas pruebas piloto sólo participan algunas de las organizaciones o entidades interesadas en el proyecto del buscador federado.
- (15) *Universia.net* se impulsa actualmente en diez países (Argentina, Brasil, Colombia, Chile, España, México, Perú, Portugal,

Puerto Rico y Venezuela), con 976 universidades ya adheridas mediante convenio como accionistas del portal que se desarrolla en cada país, para favorecer la difusión de la información universitaria, el desarrollo de las nuevas tecnologías aplicadas a la educación, la innovación educativa y tecnológica y las plataformas de comunicación educativa e interuniversitaria en el ámbito global iberoamericano. Este proyecto se concibe como el más ambicioso portal del mundo universitario hispánico. Universia.net nace con la vocación y ambición de liderar el desarrollo de la sociedad de la información universitaria, fomentando la participación de sus universidades y el desarrollo de sus proyectos, apoyando la introducción de las nuevas tecnologías, acelerando la implantación de proyectos claves, potenciando la competitividad y calidad del sistema educativo superior a través de las propuestas emanadas en el ámbito de las propias instituciones universitarias y propiciando el cambio cultural hacia la mencionada sociedad de la información. http://www.universia.net/proyectouniversia/Universidades_proyectouniversia.htm

- (16) SEO: Search Engine Optimization.
- (17) Según un reciente estudio sobre los usuarios de motor de búsqueda en España, el 92 por ciento de los usuarios utiliza Google; el 97 por ciento dice buscar contenidos en castellano (prefiere contenidos en castellano, aunque hable inglés); el 91 por ciento afirma estar buscando algo de su entorno geográfico o cultural, es decir, de nuestro propio país (un cine, un personaje de la literatura española...), no algo de fuera; y para casi un 60 por ciento, ese entorno geográfico se inscribe dentro de la propia Comunidad Autónoma del usuario del motor de búsqueda.
- (18) El problema de la web invisible tiene su origen en la forma habitual de proceder de los buscadores. Los motores de búsqueda disponen de software adecuado (*lanzan sus arañas*) para rastrear los enlaces de una página web, reconocer qué tipos de documentos hay al final de cada enlace, extraer el contenido de esos documentos y enviarlo a una base de datos global del motor de búsqueda (así es como funciona su *inteligencia*). Si la *araña* se encuentra con una *puerta cerrada* y no puede *abrirla*, tampoco puede acceder a los documentos que haya detrás: abandona su misión, deja los posibles contenidos sin explorar y, por tanto, también sin indizar. Eso es la web invisible: el contenido existe, pero los buscadores (sus *arañas* de rastreo) no pueden acceder a él ni, por tanto, tampoco transmitir o indicar su existencia y ubicación al usuario del buscador. Es como si esa información no existiera. La web permanece invisible u oculta para los motores de

búsqueda y, en la misma medida y en definitiva, para el usuario de la Red.

- (19) En el momento de redactar estas líneas, se calcula que hay en la Red unos 2.800 millones de páginas web en castellano. De ellas, Google tiene unos 250 millones. ¿De qué sirven los otros 2.300 millones de páginas, si no están en Google (responsable del 92 por ciento de las búsquedas de los hispanohablantes), si no se pueden encontrar, si no se conoce su existencia?
- (20) Puede verse al respecto: Angel Martín Municio (Director): «*El valor económico de la lengua española*», Espasa Calpe, Madrid, 2003.
- (21) Los gobiernos de Francia y Alemania decidieron, en abril de 2005, apoyar el desarrollo del motor de búsqueda «Quaero», considerado por muchos como un competidor potencial de Google. «Quaero» –que significa «yo busco»– permitiría a los usuarios buscar contenido de texto, imagen, vídeo y audio, así como también proporcionar acceso a las bibliotecas digitales europeas. El proyecto Quaero, actualmente dirigido por el grupo de trabajo franco-alemán sobre 'cooperación económica', implica también a una serie de organizaciones públicas y privadas tales como LIMSI-CNRS, Universidad de Karlsruhe, MIG-INRA, RWTH Aachen, Berlin Technologies, Jouve, INRIA, Exalead, INA, Thomson, France Telecom, Deutsche Telekom y la nueva Agencia francesa para la Innovación Industrial (Aii), creada por el presidente Jacques Chirac en agosto de 2005 y que invertirá 90 millones de euros en el proyecto Quaero. Dicho proyecto se enmarca dentro del contexto más amplio de *La Biblioteca Europea* (The European Library – TEL project), que proporciona acceso *online* a 11 millones de registros de las colecciones de 9 bibliotecas nacionales europeas. *La Biblioteca Europea* es el resultado del proyecto TEL, una iniciativa financiada por la UE para estudiar la factibilidad de un servicio pan-europeo que, finalmente, daría acceso a los recursos conjuntos de las bibliotecas nacionales de Europa. En el futuro, *La Biblioteca Europea* puede sentar las bases de una «Biblioteca Digital Europea» gigante, cuya creación fue pedida, en abril de 2005, por los dirigentes de Francia, Italia, España, Polonia y Hungría al Presidente de la Comisión Europea, José Manuel Durao Barroso, y al entonces Presidente del Consejo Europeo, Jean-Claude Juncker. Esta petición recibió el apoyo de las bibliotecas nacionales de 19 Estados Miembros de la UE. <http://europa.eu.int/idabc/en/document/4648/330> <http://ec.europa.eu/idabc/en/document/5554/194> <http://www.aii.fr/srt/aai/home>