

El español y los buscadores

23/09/2006 - Autor: José Antonio Millán - Fuente: El País

¿En qué se parece la receta del gazpacho al horario del Museo del Prado y a la vida de Cervantes? En que para conocerlos acudimos a Internet, a través de un buscador. El procedimiento funciona tan razonablemente bien que cuesta imaginar la época en que no existía. Los buscadores son la vía privilegiada de acceso a las webs, hasta el extremo de que para muchos han suplantado a los sitios: si preguntamos a un estudiante la procedencia de un dato, probablemente responda: "¡De Google!".

Los buscadores segmentan las consultas por lenguas y secundariamente por países. Los españoles son usuarios mayoritariamente de Google.es, y los franceses de Google.fr. Igual que su lengua, uno lleva consigo su buscador, y los alumnos del Liceo Francés de Madrid usan también Google.fr. En el mundo hay unas 5.000 lenguas, pero no todas se escriben, y de las que se escriben no todas tienen presencia en Internet, donde hay unas 600 lenguas. Por ejemplo, la enciclopedia colaborativa Wikipedia está disponible en 230. Los buscadores han comenzado por las lenguas con más navegantes: Google detecta 35 idiomas (aunque permite acceder a páginas escritas en otros). De las lenguas con alfabeto latino, el inglés es la más presente en la Red (casi con la mitad de páginas); en orden decreciente le siguen el alemán, el francés y el español.

Nuestra lengua abarca muchos países, pero en la web constituye un espacio único accesible a los buscadores. Google, por ejemplo, no indiza todas las páginas de la web en español, sino quizá un 40% (no el 8% que hace poco declaraba el anterior ministro de Industria). Ningún buscador recoge todas las páginas en una lengua: indizan sólo la parte más importante de su web.

Pero el uso de los buscadores no es sólo una cuestión tecnológica, sino también cultural y política. Hace año y medio Jean-Nöel Jeanneney, director de la Bibliothèque Nationale de France, publicaba un artículo en Le Monde señalando el peligro que corría la cultura europea accediendo a sus contenidos mayoritariamente a través de Google, ¡una empresa americana, y con fines de lucro! También clamaba contra el plan de digitalización masiva de libros que planea el buscador. La respuesta de la Unión Europea ha sido la futura Biblioteca Digital Europea, que estará organizada, según Jeanneney, "contra la tentación del desbarajuste, según los principios de la razón", y paralelamente se prepara un buscador franco-alemán, de nombre latino: Quaero, "busco". Presuntamente va a arreglar el también presunto sesgo norteamericano de la cultura europea en los buscadores, aunque no sabemos cómo.

¿Es ésta la mejor solución? Por ejemplo: ¿sería bueno un nuevo buscador de Internet para el español? Depende: la tecnología de búsqueda lleva años de desarrollo y no parece fácil emularla. Indizar mayor porcentaje de la Red en español puede tener unos costes

exorbitantes... para rendir un servicio muy poco mejor. Primar resultados españoles o hispanoamericanos frente a los de otros orígenes (para así defender una identidad cultural) es absurdo, si lo que se quiere es dar un buen servicio. Incluir bases de datos ahora opacas a los buscadores puede ser útil, aunque sus instituciones responsables también pueden hacer acuerdos para integrarlas en los buscadores actuales.

Aunque caben mejoras locales. Una empresa de Telefónica ha lanzado el buscador Noxtrum (como se ve, el latín está de moda), que usa las bases de datos comerciales de las Páginas Amarillas, mientras que el Instituto Cervantes anuncia un Buscador Panhispánico (que ojalá cambie un nombre de resonancias tan imperialistas), con catalogación de contenidos y una selección de sitios hecha por expertos.

Es dudoso que un nuevo buscador compense el esfuerzo que costaría, pero hay cosas que se pueden hacer a favor de la web en español. La primera: crear más y mejores páginas en nuestra lengua. Ésta es una tarea para instituciones como bibliotecas (que ya están incorporando a la Red nuestro patrimonio cultural) o universidades (que comienzan a publicar digitalmente su producción), pero también para las empresas y los ciudadanos. Por ejemplo: un estudio reciente demuestra que las editoriales hacen escaso uso de la web para promocionar sus libros; los blogs profesionales o la Wikipedia española podrían crecer y mejorar.

El segundo requisito es que las páginas faciliten el trabajo de los motores de búsqueda, haciendo que sus contenidos sean accesibles sin barreras para cualquier buscador (en vez de ahogar sus datos en animaciones). Así, hoy podrá indizarlas Yahoo o Google, y mañana, ¿quién sabe? En búsqueda estamos ante un oligopolio de facto, pero la mejor manera de evitarlo es permitir que puedan competir otros agentes.

Pero hay un flanco en el que podríamos intervenir: las tecnologías lingüísticas. Cuando lleguen los buscadores "semánticos", será imprescindible que haya sistemas automáticos para la lengua española. Un ejemplo simple: ante la demanda de un "pintor en Córdoba" el buscador podría preguntar "¿de paredes o de cuadros?", y al analizar las páginas el sistema debería deducir que el pintor de un documento que habla de brochas y rodillos es del primer tipo, y el que está con óleo o retrato del segundo. Crear programas que "entiendan" el español es un objetivo que deberían emprender los investigadores y las empresas del mundo hispanohablante... o nos vendrán de fuera.

El simple ejercicio de pensar qué tipo de buscador necesita el español pone en cuestión nuestro modelo de desarrollo tecnológico y de política lingüística. Las empresas que dominan Internet nacieron en garajes o universidades americanas, consiguieron capital riesgo y dominaron el mercado. Hagamos que avance la investigación en nuestras universidades (que generan poquísimas patentes), que florezcan pequeñas empresas (ahora envueltas en trabas burocráticas) y que puedan captar capitales, o no podremos quejarnos cuando otros hagan lo que nosotros no hicimos. Lo que necesita Europa no es repetir lo que ya existe (so pretexto de que no refleja nuestra cultura), sino vivir el espíritu de innovación de donde sale lo que existirá mañana.

José Antonio Millán es autor de Internet y el español.

Webislam