

# **SENCILLEZ Y NATURALIDAD EN EL ACCESO A LA INFORMACIÓN: LA BIBLIOTECA DIGITAL DEL INAP**

**Josu Gómez Pérez**  
Director de Desarrollo  
Bitext.com

**Enrique Torrejón**  
Director de I+D  
Bitext.com

## **Palabras clave**

*Accesibilidad, brecha digital, lenguaje natural, administración próxima, biblioteca digital*

## **Resumen de su Comunicación**

*El siglo XXI comienza a recibir el nombre de "Siglo de la información". Cada vez es más importante poder ofrecer a los ciudadanos puntos únicos de acceso a las colecciones documentales, que permitan extraer el máximo rendimiento a la información actualizada de una institución. Un ejemplo es la nueva Biblioteca Digital que ofrece el Instituto Nacional de Administración Pública (INAP): un punto de acceso a distintas fuentes de información, externas e internas, accesible dentro de la página web del INAP. Permite localizar información entre los documentos en formato digital disponibles por la Biblioteca del INAP y acceder al contenido de dichos documentos desde un entorno de uso sencillo e intuitivo, utilizando las técnicas más avanzadas de tecnología lingüística para asegurar la recuperación óptima de la información, como por ejemplo interfaces en lenguaje natural.*

*Con este instrumento la Biblioteca del INAP pone a disposición de todos los ciudadanos su rica información documental, superando los paradigmas de acceso ligados a una ubicación geográfica, y ofreciendo además un acceso natural a la documentación, permitiendo realizar las consultas con las propias palabras del usuario, haciendo así este acceso sencillo y efectivo independientemente del nivel técnico o la edad de cada usuario. En resumen, la Biblioteca Digital rompe con las limitaciones de espacio y de acceso, universalizando de esta forma el acceso a la información.*

# **Sencillez y naturalidad en el acceso a la información: la Biblioteca Digital del INAP**

## **Índice:**

- **La accesibilidad en la Administración Pública: hacia una Administración próxima**
- **El caso del INAP: gran cantidad de información con accesibilidad limitada**
- **La Biblioteca Digital como punto de acceso integrado a la información**
- **La importancia del lenguaje natural para acceder a la información textual**
- **Conclusión: hacia una Administración que hable el idioma de los ciudadanos**

## **La accesibilidad en la Administración Pública: hacia una Administración próxima**

Las Administraciones Públicas están potenciando el uso de Internet como canal de comunicación con el ciudadano, y con mucho acierto, hasta el punto de que la administración electrónica puede producir una revolución en la relación ciudadanos/administradores, mejorando servicios y reduciendo costes. Probablemente, el mayor obstáculo para esta revolución puede ser el lenguaje: ¿se va a comunicar el ciudadano con la administración en "indio"?

Cuando buscamos información en Internet, ya nos resulta familiar a muchos utilizar expresiones como: "pagar impuestos coche". Y también nos resulta igualmente familiar obtener como respuesta un documento que no está relacionado con nuestra consulta, como: "Desarticulada red defraudó 54 millones al importar coches sin pagar impuestos. Toledo, 20 ago (EFE).-".

Estas situaciones son habituales para los que usamos buscadores para localizar información en Internet. Y la que acabamos de describir es real. En otras palabras, nos hemos acostumbrado a «hablar en indio», como en las películas del Oeste (en inglés, "keywordese", derivado de "keyword", "palabra clave"). A pesar de que se puedan producir estas situaciones, los buscadores son actualmente, con diferencia, la mejor herramienta para localizar información. Gracias a ellos podemos sacar partido a todos los contenidos que alberga Internet.

De hecho, probablemente, la búsqueda anterior acabe en que descubramos que el nombre habitual de este impuesto es "impuesto de circulación", y a partir de ahí prosigamos la búsqueda hasta dar con lo que buscamos.

La mejor prueba de que este proceso es así es que todos los que buscamos

habitualmente en Internet lo vemos como un proceso iterativo, repetitivo; un proceso de ensayo-error. Éste es el proceso típico de búsqueda para una parte de la población, una parte pequeña, la que usa Internet y los buscadores a diario, casi siempre como herramienta de trabajo.

El resto de la población, la inmensa mayoría, tendría más problemas para culminar con éxito la búsqueda anterior. Esta inmensa mayoría sigue preguntando, como hemos hecho siempre, cosas como "puedo pagar el impuesto del coche por Internet", "quiero pagar el impuesto de mi coche". Esta inmensa mayoría es más que significativa pues supone alrededor del 90% de la población: según el informe "La sociedad de la información en España 2006" de Fundación Telefónica, sólo el 10% de la población española hace un "uso estable" de Internet.

Dado el éxito que Internet ha tenido como herramienta barata y efectiva para distribuir información (ahora también para compartirla y crearla), muchas instituciones se plantean desde hace ya un tiempo usar Internet para dar servicios de información al ciudadano en general. Las Administraciones Públicas, en todos sus niveles, están ya aprovechando todas las posibilidades que ofrece la Red para comunicarse con el ciudadano. La idea es tan acertada como prometedora (y es muy acertada): Internet puede ser, y lo es en muchos casos, un excelente canal para facilitar, potenciar y flexibilizar las relaciones entre Administración y ciudadanía.

La Administración Electrónica, también llamada "Administración Próxima", puede ser una forma mejor de administración, más democrática y efectiva, no sólo un complemento de la Administración en su forma actual.

Sin embargo, en nuestra opinión, uno de los obstáculos más importantes para alcanzar este objetivo es precisamente la forma en que los buscadores funcionan ahora; tal y como hemos ejemplificado al comienzo, el experto en buscadores debe saber buscar en "indio". ¿Por qué? Los ciudadanos a los que se dirige la Administración no hablan en indio ("pagar seguro coche"), hablan el castellano (o el catalán, valenciano, gallego, euskera...) que han hablado siempre y que han usado en muchos otros canales (teléfono, carta, fax, entrevista presencial, correo electrónico...). La inmensa mayoría de los ciudadanos sabe decir: "quiero pagar el impuesto de mi coche", pero no sabe decir "pagar impuesto coche"; es decir, sabe usar el lenguaje natural, el español, pero no sabe expresarse en indio, ni tampoco sabe que éste es el método más efectivo para dar con lo que buscamos.

Estos problemas se presentan con una virulencia mayor cuanto mayor es la cantidad de información que un organismo posee, y cuanto más interés tiene en hacer esa información accesible a una gran cantidad de personas, no todas ellas avezadas técnicamente.

### **El caso del INAP: gran cantidad de información con accesibilidad limitada**

El INAP (Instituto Nacional de Administración Pública), como define su

página web, es el centro de la Administración General del Estado responsable de la selección y formación de sus directivos y empleados públicos, e impulsa tareas de investigación y estudio al servicio del proceso de modernización de las Administraciones Públicas.

Estos fines requieren que el organismo tenga disponible en todo momento y accesible para una gran cantidad de personas (todos los empleados públicos del Estado) volúmenes importantes de información y documentación. Para cumplir sus objetivos fundacionales el INAP debe recopilar, almacenar, clasificar y hacer accesibles un gran número de materiales de divulgación y de investigación a todas las personas que se puedan interesar por ellos. Además, como servicio público, esta información tiene que estar lo menos restringida posible.

Uno de los servicios más importantes del INAP es su Biblioteca. Se trata de un centro bibliográfico y documental especializado en Administración pública, que constituye en la actualidad la colección más importante sobre estas materias de cuantas existen en España y una de las más relevantes a nivel internacional. Su catálogo bibliográfico, de libre acceso, está especializado en Ciencia Política, Derecho Administrativo, Derecho Constitucional, Administración y Gestión Pública, Administración Local y Autonómica, Función Pública, Hacienda y Economía, Derecho Privado, Medio Ambiente, Ordenación del Territorio y Urbanismo, Historia de las Administraciones, Sociología...

Además, la Biblioteca es un lugar de acceso preferente a toda la extensa labor de publicación que el Servicio de Publicaciones del INAP lleva a cabo, por medio de monografías y de publicaciones periódicas que el mismo organismo edita.

Por los propios objetivos fundacionales del INAP se intenta que el acceso a los materiales de la Biblioteca sea lo más abierto posible. No obstante, el hecho de tratarse de una biblioteca física, que debe ser asistida en todo momento por empleados y para la cual se requiere un acceso presencial, impone una serie de limitaciones (de horario, de libertad de acceso, de préstamo...) que condicionan en cierta medida su uso.

### **La Biblioteca Digital como punto de acceso integrado a la información**

Para mejorar la accesibilidad a la información contenida en su biblioteca, el INAP puso en marcha desde su página web el proyecto que ha visto la luz pública recientemente: su Biblioteca Digital.

La Biblioteca Digital del INAP es un punto de acceso a distintas fuentes de información, externas e internas, accesible dentro de la página web del Instituto Nacional de Administración Pública. Permite localizar información entre los documentos en formato digital disponibles por la Biblioteca del INAP y acceder al contenido de dichos documentos desde un entorno de uso sencillo e intuitivo, utilizando las técnicas más avanzadas de tecnología lingüística para asegurar la recuperación óptima de la información.

La información disponible por medio de este servicio se puede dividir en

materiales publicados (Monografías, Publicaciones Periódicas y Otras), páginas web externas al propio INAP pero con información relevante, catálogos bibliográficos propios y bases de datos a las que el centro está suscrito, entre otras fuentes de información.

Además, la Biblioteca Digital permite realizar, desde la pantalla principal, búsquedas globales en todas las fuentes de información a la vez, ofreciendo después la posibilidad de desglosar los resultados según el tipo de fuente.

Ofrecer un punto de acceso sencillo para todos los usuarios (desde los investigadores más tecnificados hasta los menos acostumbrados a la tecnología) requiere, además, que la recuperación de la información y la selección de las fuentes sea lo más intuitiva posible. Para cumplir este objetivo, la Biblioteca Digital integra las soluciones de lenguaje natural de Bitext, que permiten a los usuarios consultar los distintos repositorios utilizando su propio lenguaje.

### **La importancia del lenguaje natural para acceder a la información textual**

Acceder a la información textual disponible en los sitios web de la Administración Pública usando "indio" carece, por un lado, de la capacidad expresiva que aporta el lenguaje natural e imposibilita, por otro lado, el acceso a los contenidos de la Administración a la inmensa mayoría de la ciudadanía que usa lenguaje natural.

El tratamiento computacional del lenguaje natural es la clave en este aspecto para acceder adecuadamente a la información textual contenida. Y dicho tratamiento va más allá del clásico stemming de los buscadores: se trata de saber si las palabras están escritas correctamente, de conocer la morfología de las palabras, de ver la estructura sintáctica de las consultas que hacen los usuarios y de extraer el significado de las mismas para, a continuación, buscar de forma efectiva en los contenidos.

De esta manera, la Biblioteca Digital cuenta con herramientas que permiten hacer mucho más intuitiva la experiencia de recuperación de información de los usuarios, como:

#### 1) Corrección ortográfica

Antes de que el usuario lance su consulta, todo buscador debe ser capaz de comprobar que las palabras de dicha consulta tienen la ortografía correcta pues, si no es así, la búsqueda será infructuosa. La importancia de un corrector ortográfico integrado en los buscadores que atienden las consultas de los ciudadanos no se puede menospreciar. Si un ciudadano introdujera una consulta como "deseredar a mis hijos" en cualquier buscador de la Administración Pública, muy probablemente no encontrará ninguna información al respecto debido a la falta ortográfica en el verbo. La Biblioteca Digital identifica estos errores ortográficos y devuelve sugerencias al usuario, como

## **deseredar -> desheredar**

### 2) Comprensión lingüística de la consulta

Un buscador que entienda el lenguaje natural que usan los ciudadanos ha de ser capaz de conocer las palabras dentro de la frase, no tratarlas como un mero conjunto de caracteres. Así, si algún ciudadano pregunta por "cómo desheredar a mis hijos", el buscador debería ser capaz de recuperar documentos donde se hable, por ejemplo, de "hijos justamente desheredados". La Biblioteca Digital incorpora este tipo de tecnología, devolviendo los resultados correctos aunque el texto no coincida exactamente con los términos de la consulta, como

#### **desheredar -> desheredados**

### 3) Procesamiento de fenómenos complejos

Además de la variabilidad morfológica, un buscador que entienda el lenguaje natural tiene que ser capaz de analizar sintácticamente las consultas de los ciudadanos. Tiene que saber manejar, por ejemplo, la coordinación copulativa y disyuntiva, o la negación. Así, no es lo mismo buscar "desheredar a mis hijos y cónyuge" que "desheredar a mis hijos pero no a mi cónyuge". Un buscador que devuelva los mismos documentos para las dos consultas anteriores no es un buscador que puede atender a las necesidades de información del ciudadano. La Biblioteca Digital identifica este tipo de estructuras, como

#### **leyes de administración local pero no de España**

### **Conclusión: hacia una Administración que hable el idioma de los ciudadanos**

Todas estas peculiaridades deben ser tenidas en cuenta a la hora de ofrecer un acceso a la información digital, si se quiere que esta información no sea sólo accesible a un número pequeño de personas con gran habilidad tecnológica (lo que se ha dado en llamar la "brecha digital"). Con esta situación en la mente, el INAP no se ha limitado a poner a disposición de sus usuarios su información digital, sino que ha facilitado, por medio de herramientas de lenguaje natural, el acceso intuitivo a esta información.

El INAP, por medio de su Biblioteca Digital, está marcando el camino hacia la plena accesibilidad de los contenidos digitales, en un promisorio paso para conseguir los objetivos que el reto de la Administración Próxima ha marcado.