

# El sistema de localización de recursos de información “Infojoven.com”

**José Juste Jordán**

**Lorena Monge Vázquez**

**Daniel Pons Betrián**

Instituto Aragonés de la Juventud (España)

## 0.1. Resumen

“Infojoven.com” es un proyecto de cooperación, intercambio y difusión selectiva de información juvenil. El sistema pretende desarrollar una red de trabajo horizontal, libre y voluntaria que permita compartir contenidos y crear herramientas de explotación del conocimiento. La información recogida en Infojoven sigue los principios de la “carta Eryca”. Los usuarios del sistema son los profesionales de la información juvenil que suministran información específica a los jóvenes, los usuarios finales, optimizando recursos y utilizando información validada y verificada. El sistema de información sigue una arquitectura cliente-servidor, que evoluciona hacia cliente-*broker* con el objetivo de optimizar el uso de la red. Una vez que la información ha sido recuperada, el sistema contiene un módulo para distribuirla mediante una central de comunicaciones.

**Palabras clave:** Información juvenil. Sistemas de información. Recursos de información. Instituto Aragonés de la Juventud.

## 0.2. Abstract

“Infojoven.com” is a project of cooperation, exchange and selective dissemination of youth information. The system pretends to develop a free and voluntary horizontal network that will allow sharing contents and providing tools of exploitation. The data contained in “Infojoven.com” follows the principles of the “Eryca letter”. Youth Information professionals will work with “Infojoven.com” to provide specific information to young people, optimizing resources and using validated and verified information. The system has a client-server architecture that evolves toward a client-broker one to improve the use of the net resources. The system contains a module used for disseminating the information that has been obtained through Internet and other electronic resources.

**Keywords:** Youth information. Information systems. Aragon (Spain).

## 1. Introducción

Infojuven.com (2003) es un proyecto promovido por el Instituto Aragonés de la Juventud (IAJ), que desarrolla una red de trabajo en común entre los profesionales de la información juvenil que se enmarca en la nueva política de comarcalización del Gobierno de Aragón. Es un Sistema de Localización de Recursos de Información (SLRI), que tiene por objetivo desarrollar una red horizontal libre y voluntaria, una experiencia modelo de gestión de la información adaptada al marco de cooperación europea, que permita compartir contenidos y proporcionar herramientas de explotación del conocimiento generado por estos servicios y para ellos. Es una nueva forma de hacer y ver la información para jóvenes adaptada a la nueva realidad administrativa de Aragón, basada en una fuerte descentralización. Tecnológicamente, se inserta en la corrientes del software libre, promoviendo el desarrollo de tecnologías propias de libre distribución, que permitan compartir información y generar conocimiento. En la misma línea surge la idea de "Documentación Libre", entendida como documentación que puede ser accedida libremente. Actualmente vivimos en la Sociedad de la Información; la información obtiene valor al compartirla y distribuirla, no por su posesión. Éstos son los pilares básicos de nuestro sistema.

Se utilizan estándares abiertos para trabajar con protocolos de comunicación, formatos de almacenamiento de datos, etc. De esta forma, se favorece la interoperabilidad y compatibilidad con otros sistemas, garantizando futuras ampliaciones y mejoras. Utilizamos el formato RTF para los textos, y XML para las respuestas a las búsquedas en Internet, consultas a bases de datos en SQL, etc.; evitando utilizar en la medida de lo posible extensiones privativas de programas.

Es un proyecto que se ha desarrollado por los propios jóvenes, mediante la realización de prácticas laborales en las instalaciones del IAJ, la realización de un proyecto de fin de carrera, y un convenio con la Universidad de Zaragoza a través de Universa y del Departamento de Educación y Cultura del Gobierno de Aragón. No es un proyecto académico de investigación, sino que es un proyecto plenamente operativo en la red de centros de información juvenil de Aragón. Es un trabajo en equipo entre documentalistas e informáticos. Para llevar a cabo este proyecto y convertirlo en un sistema de difusión de información, se ha hecho hincapié de igual manera durante su fase de creación en el aspecto de soporte técnico y en el documental, intentando unir una buena metodología tanto en la recuperación como en la calidad y fiabilidad de la información.

## 2. Objetivos

El proyecto va encaminado a optimizar los servicios que prestan los técnicos juveniles en materia de información juvenil. Cualquier joven que necesite res-

puestas en temas relacionados con el acceso a la vivienda, convocatorias, cursos, orientación laboral, etc., podrá requerir al técnico juvenil, que a su vez hará uso del sistema Infojuven para satisfacer las necesidades de información solicitadas. Los jóvenes pueden realizar consultas y a su vez participar en la creación de nueva información.

El objetivo principal es propiciar la comunicación y el intercambio, siguiendo el concepto de *clearing house*: actuación en red junto con otros centros en régimen de ayuda mutua. La finalidad es que la información disponible llegue selectivamente a los interesados siendo difundida y distribuida entre las personas que la precisan. Sus objetivos específicos son ofrecer abiertamente a los usuarios la máxima información actualizada y de calidad; y promover y aplicar reglas, normas y sistemas establecidos para facilitar el intercambio de información. Centra su trabajo en la descripción del contenido, en el tratamiento y la difusión del mismo. Su misión consiste en identificar con la mayor precisión las informaciones que pueden ser útiles a los usuarios. Para conseguir esta misión se ha de sustentar en cuatro principios: seleccionar, tratar, difundir y dinamizar la documentación.

### 3. Usuarios

Infojuven.com esta planteado como una fuente de recursos (figura 1) para profesionales de la información juvenil —usuarios directos—, que realizan funciones de intermediación con los jóvenes —usuarios finales—. Tiene como objetivo que los profesionales participantes en el proyecto adquieran conocimiento en el manejo de herramientas estándar en el campo de la documentación,



Figura 1. Página web de explotación de Infojuven

que les permitan desarrollar mejor su trabajo cotidiano y afrontar la utilización de Internet de la manera más profesional; optimizando todos sus recursos. De esta forma, se puede compartir la producción informativa haciendo uso de una información validada y comprobada con otros profesionales de la información juvenil. Potencia y dignifica el trabajo de los técnicos de juventud de las comarcas aragonesas, al posibilitar su trabajo como asesores o consultores de los jóvenes, y no como meros informadores juveniles.

#### **4. Proceso de información**

El sistema se estructura en torno al modelo clásico del ciclo de la información —recogida, tratamiento, difusión—, en el que la recuperación de información está presente en cada una de las fases del tratamiento documental.

##### **4.1. Recogida de la Información**

La información que se selecciona para su difusión se centra en los temas de interés para los Servicios de Información Juvenil, siguiendo los criterios de un manual de estilo (Colás, Juste y Monzó, 1998) y teniendo como principio la carta ERYICA (2003), adoptada en Bratislava (Eslovaquia) el 3 de diciembre 1993 por la 4.<sup>a</sup> Asamblea General de la Agencia Europea para la Información y el Asesoramiento de los Jóvenes (ERYICA). Establece entre otras cosas la no selección de noticias con contenido político y que puedan ser objeto de crítica a las líneas políticas de los Gobiernos, no sexistas, no racistas, no discriminatorias, etc. En su redacción, la noción de “servicio de información juvenil” engloba a los servicios que se asocian al acto de informar —el asesoramiento y el apoyo— y que contribuyen a garantizar el derecho de los jóvenes a la información.

Las materias que en la actualidad aborda y comparte infojuven.com son accedidas desde un menú principal son las siguientes: dirección de páginas web, directorio físico de entidades, noticias de interés para los jóvenes, eventos de actualidad, legislación juvenil, reseñas de prensa, convocatorias, ofertas de empleo, reseñas gráficas de actividades, soporte informático, buscadores y portales sobre bases de datos, días oficiales o señalados, fiestas patronales, ferias, cartelera de espectáculos, directorio de centros de información, documentos de congresos, informes no publicados, revistas para jóvenes, foro de consulta especializado por materias, y preguntas frecuentes. Da una gran importancia a la información no formal que se adquiere a través del foro. El sistema pretende girar en torno a este servicio, con el objeto de capturar y crear un gran banco de datos de información no formal que habitualmente se pierde.

Una vez se ha realizado la recopilación y selección de información que se va a añadir al sistema, se procede a la introducción de datos. Este proceso se realiza desde un conjunto de programas, desarrollados con el lenguaje de programación

Delphi, que funcionan en un sistema operativo Windows 98 o superior. Esta elección se ha realizado para ajustarse a las máquinas sobre las que se trabaja en los centros de información juvenil, que en la actualidad sólo disponen de éste sistema operativo. Se dispone de dos niveles de permiso de acceso al sistema. Por un lado, se puede acceder en modo sólo lectura, de tal forma que sea posible consultar toda la información. Este es el método utilizado para el acceso público a través de Internet. Por otro lado, están los centros de información juvenil que participan en el proyecto, a los cuales se les provee de un nombre de usuario y una contraseña privados para conectarse al sistema, momento a partir del cual dispondrán la capacidad de añadir, modificar y eliminar la información almacenada, ingresando en una red distribuida de iguales. El proceso de introducción de información se lleva a cabo desde sistemas conectados a Internet, ya sea por tecnologías de comunicación GPRS, UMTS, RTC o banda ancha, que permitan la conexión por el protocolo de Internet (TCP/IP).

Hay que tener en cuenta que potencialmente puede haber un gran número de centros que dispongan de permiso de modificación, y a su vez, otro gran número de personas que contribuyan por cada centro. El sistema se basa en la confianza, sin un filtrado previo a la hora de introducción de información. Esto plantea problemas de contenidos, ya que la información puede ser modificada por cualquiera. Para afrontar este problema, existe una supervisión a dos niveles: por un lado, una supervisión colegiada, que son sugerencias realizadas por los centros con acceso a modificación de datos; y, por otro lado, hay una supervisión técnica, realizada por una persona concreta, responsable de asistir a los centros proporcionando un aprendizaje continuo. La supervisión se realiza a partir de un muestreo de la información.

La información es tratada desde el nivel de dato y no como documento. Se han desarrollado tecnologías propias y específicas que simplifican el proceso de transformación de la información en datos. Estos datos se clasifican de acuerdo a una clasificación decimal temática propia, la Clasificación Decimal Juvenil, creada, consensuada y utilizada por los Sistemas de Información Juvenil (SIJ). La clasificación utilizada en la descripción de la información (figura 2) se complementa con auxiliares de lugar y forma, que ayudarán a restringir los resultados de las búsquedas. La principal virtud de esta clasificación es que es el único factor de conexión entre todos los centros, al ser utilizada y conocida por la totalidad de los profesionales de la información juvenil. Esta fuente de datos, se almacenará tras su análisis en una base de datos sobre Internet, obteniendo una estructuración pura de la información compartida y accesible desde cualquier lugar.

## 4.2. Tratamiento de la Información

El eje principal del sistema es un motor informático que permite optimizar y mejorar el proceso de la recuperación de la información. Su tecnología ha sido completamente desarrollada *ad hoc* para este proyecto. Este motor actúa como una capa transparente entre el programa cliente y el servidor MySQL, haciendo las funciones de *broker*. Realiza un conjunto de tareas, tanto técnicas como documentales: tratamiento léxico, tratamiento semántico, búsqueda por clasificación y sus auxiliares, y sistema de caché para acelerar la respuesta. Esta evolución permite múltiples ventajas en el servicio, ya que el *broker* actúa como intermediario en las búsquedas eliminando tráfico en entornos saturados (Internet). El motor principal utiliza internamente un sistema de caché, diccionarios y técnicas documentales para optimizar la velocidad y la calidad de las respuestas.

Las técnicas de trabajo sobre este servidor se basan en el lenguaje estándar de interrogación a bases de datos, SQL (Groff y Weinberg, 1998), con el objetivo de utilizar un sistema plano de consultas, estándar, abierto y común, que relaciona múltiples tablas y además permite una amplia versatilidad de manejo. Se han añadido unas extensiones auxiliares propias al lenguaje para incrementar la funcionalidad, al objeto de garantizar la portabilidad y adaptarse a las necesidades según los requerimientos de uso. Por su lenguaje plano altamente simplificado reduce los costes de aprendizaje (permite realizar por parte de no informáticos consultas complejas y muy personalizadas). Los nuevos portales se están desarrollando

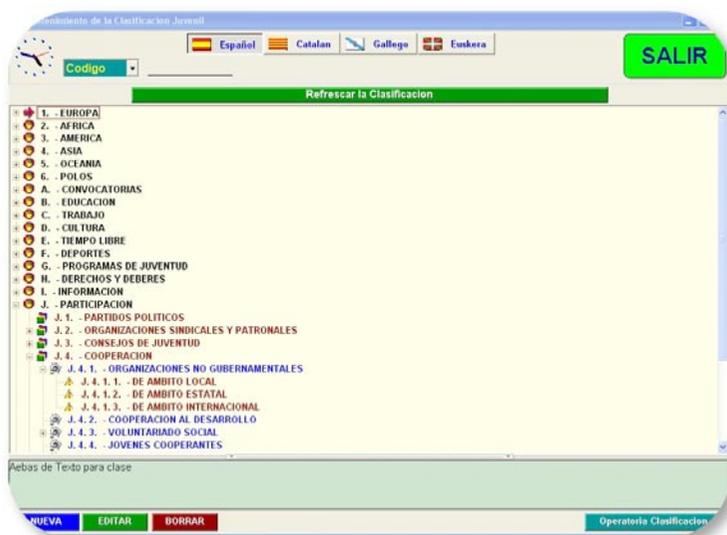


Figura 2. Clasificación Decimal Juvenil

sobre bases de datos, lo que les aporta un gran dinamismo y enorme ligereza en el mantenimiento, pero los hace opacos a los buscadores tradicionales, ya que el contenido de las páginas puede cambiar en cada acceso y para poder localizar la información es necesario bucear en los contenidos por medio de interrogaciones. Así mismo, la personalización que permiten a los usuarios hace que los buscadores tradicionales se encuentren con grandes dificultades a la hora de indexar sus contenidos. Es necesario desarrollar nuevas técnicas de búsqueda que localicen la información dentro de estas bases de datos, para lo cual es necesario desarrollar nuevas estrategias que permitan establecer vínculos de confianza por medio de protocolos de consulta, que además de ser eficaces no comprometan la seguridad de los contenidos, muy sensibles en el caso de las administraciones públicas.

### 4.3. Difusión de la Información

Tras la recuperación de los datos almacenados, la información se personaliza y diseña de acuerdo a la imagen corporativa de cada centro comarcal, proporcionando para ello las herramientas necesarias — editores de textos, tabulares o gráficas) que, accediendo a la información compartida, puedan generar formatos propios. Para la distribución personalizada de la información y la elaboración de dossiers, existe una tecnología propia de computación distribuida, que permite aprovechar los equipos informáticos de los centros cuando no se están utilizando, realizando de forma cooperativa los trabajos de tratamiento y distribución de la información. Los centros de información juvenil podrán recuperar los datos mediante senten-



Figura 3. Central de comunicaciones

cias SQL, que, junto con unas plantillas de búsqueda y de formato, pueden generar boletines, informes o dossieres. El resultado de las consultas será distribuido a los jóvenes electrónicamente mediante sistemas de difusión selectiva de información en función de sus perfiles personales y del ámbito de la entidad, utilizando para ello todas las herramientas telemáticas actualmente disponibles —fax, correo electrónico, SMS, MMS, llamada de voz sintética, FTP—. No se trata de crear sistemas automatizados, sino de crear herramientas que automaticen el trabajo. Se potencia el desarrollo de páginas web por parte de cada centro participante al posibilitar mantener los contenidos informativos de las mismas en las tablas del sistema Infojoven con la personalización propia de cada uno de ellos, de forma totalmente automatizada por programación en PHP, Java, ASP, etc.

## 5. Sistema de búsqueda

### 5.1. Inicio

El motor de búsqueda es un sistema informático que se basa en la intermediación entre el cliente y el servidor desarrollando una arquitectura cliente-*broker*, actuando como un agente intermediario que optimiza y trabaja las consultas. Se basa en el lenguaje estándar de interrogación a bases de datos SQL accediendo a una base de datos MySQL. Cuando el cliente lanza una consulta al sistema, es convertida a una sentencia SQL que es capturada por el *broker*. A partir de aquí se empieza a trabajar con la consulta, realizándole las optimizaciones necesarias y utilizando posteriormente el sistema documental para recuperar la información.

### 5.2. Tratamiento de la consulta

En primer lugar, se procede a la eliminación de palabras vacías por medio de un diccionario de artículos, preposiciones y conjunciones que se consideran no útiles en un proceso de recuperación de información. A las palabras que terminen en ‘s’ o en ‘es’, se les quita la desinencia, y se realiza una búsqueda con el término original y con el obtenido. Se tienen en cuenta tratamientos de casos especiales, como pez-peces, en los cuales se cambia la z por una ‘c’ al formar el plural.

Las búsquedas se realizan sin distinguir mayúsculas y minúsculas. En los casos de palabras que contienen tildes, se procede a la búsqueda con tildes y sin tildes. Los caracteres que no son alfanuméricos no son tratados, y se toman como separadores. Si se introduce un nombre de mes, un formato de fecha, o un número que represente a un año, éste se busca sobre los campos de fecha. Los verbos y adjetivos no son objeto de tratamiento.

Se realiza una corrección ortográfica de los términos introducidos, utilizando un diccionario ortográfico. En los casos en que se detectan errores de ortografía, se aplican las siguientes normas: Si al buscar en el diccionario una palabra intro-

ducida, no se obtiene resultado, se considerará que la palabra es correcta, aun sin ser reconocida, y se avisa al usuario de este hecho. Si en una búsqueda se obtiene un único resultado, o bien la palabra es correcta, en cuyo caso no se hace nada; o bien la palabra no se ha encontrado en el diccionario, aunque se ha localizado una palabra muy similar, en cuyo caso se considera que se ha corregido una posible falta de ortografía. Si en la búsqueda se obtienen varios resultados, significa que la palabra no se ha encontrado, pero que existen posibles alternativas; en ese caso se busca con la palabra introducida originalmente, ofreciendo posteriormente al usuario las posibles sugerencias. Hay que destacar que todas las opciones de tratamiento léxico y de ortografía se pueden deshabilitar, permitiendo realizar búsquedas manuales, pero, por contra, menos inteligentes.

### 5.3. Realización de la búsqueda

Las búsquedas se realizan (figura 4) por patrón decimal sobre la clasificación temática. Una vez encontrado un número, se tienen en cuenta todas las que hay por debajo de éste. También se puede especificar que la búsqueda tenga en cuenta los más genéricos. Las búsquedas se realizan sobre los términos de la clasificación combinándolos con el operador AND. Si se obtienen menos números de clasificación que un umbral predeterminado, se combinan con el operador OR. Si la cantidad de números sigue estando por debajo del umbral, se procede a buscar resultados sin tener en cuenta la clasificación. Si una palabra es encontrada en la clasificación como auxiliar de forma o de lugar, ya no se seguirá utilizando para



Figura 4. Sistema de búsqueda

buscar información en el resto de la base de datos. Si una búsqueda contiene una palabra específica, tal como ‘legislación’, ‘campamentos’, etc., que está en relación con una tabla del sistema, se restringe a esa tabla. Tras la obtención de resultados, se muestra una opción de ampliación de búsqueda al resto del sistema de información.

#### **5.4. Resultados de la búsqueda**

Una vez obtenidos los resultados de una búsqueda, se pueden realizar búsquedas sobre ellos. Se está trabajando en mejorar la ordenación del listado de resultados. La finalidad es que los primeros resultados de las búsquedas sean los más apropiados e interesantes. Para ello, cada registro concreto de una tabla contiene un valor numérico que se incrementa cada vez que es consultado. Con este proceso, se consigue dar más peso a los registros más consultados, con una filosofía similar a Google, que da más importancia a las páginas más referenciadas por otras. En el proceso continuo de mejora del sistema, se utilizan las bitácoras generadas por cada búsqueda realizada. Se analizan las búsquedas introducidas, determinando en primer lugar la calidad de la sentencia de búsqueda, y posteriormente se valoran los resultados obtenidos. El técnico documentalista determina si la búsqueda ha sido sobre un tema válido, si ha sido formulada correctamente, y si los resultados obtenidos son coherentes con los esperados. Con esta información se realiza un proceso de retroalimentación de inteligencia al sistema, ampliando o depurando el proceso informático y documental utilizado en las búsquedas.

#### **5.5. Presentación de la búsqueda**

El motor de búsqueda puede devolver los resultados en formato XML. De esta forma se consigue recuperar un documento en formato estándar y abierto. El lenguaje XML es un lenguaje de marcado, que permite trabajar con documentos organizados por etiquetas personalizadas. Los documentos recuperados en XML pueden ser mostrados a través de Internet, junto con una hoja de estilo que les da formato, o en un programa editor de XML, desarrollado específicamente para poder realizar búsquedas y selecciones sobre las etiquetas que estructuran el texto XML. Este sistema nos proporciona la opción de recopilar información útil en un nuevo documento que puede alimentar de nuevo a nuestro sistema de información o a otros.

Se dispone también de un documento XSD para describir la estructura, contenido y semántica de un documento XML, mediante una descripción en el lenguaje estándar XML Schema. XML Schema es un lenguaje, escrito en XML, que surgió de la necesidad de poner los Document Type Definition (DTD) en formato XML. Está en desarrollo el soporte de RFD para definir las cabeceras HTML de la parte con salida web del sistema, de tal forma que se facilite la labor

de la clasificación de los documentos. También se plantea devolver documentos RSS para favorecer la propagación rápida de cabeceras de noticias.

## **6. Conclusión**

Infojuven aúna tecnología informática y documental desarrollando un sistema informacional que trata la información como origen de datos, almacenándolos en tablas interrelacionadas por medio de una clasificación decimal. El dato almacenado es a su vez transformado en información, mediante un conjunto de aplicaciones y técnicas documentales e informáticas, para su distribución a los usuarios finales (los jóvenes). La gestión del sistema se realiza por medio de un conjunto de aplicaciones informáticas desarrolladas a medida, que dan soporte a las necesidades del trabajo documental del sistema, por medio de asistentes de estructuración y consulta. El resultado se formatea por medio de editores especializados, que permiten generar documentos y dossiers para su distribución.

La distribución es realizada por medio de un sistema desarrollado *ad hoc*, que da soporte a todos los medios de difusión electrónica de información: fax, correo electrónico, SMS, MMS, y síntesis de voz. Las búsquedas se realizan por medio de un motor específico, que utiliza como elemento básico de sus indexaciones la clasificación decimal. Este motor da soporte tanto a las búsquedas en Internet a través de un navegador web como a los asistentes de las aplicaciones informáticas.

El sistema parte de los principios filosóficos y éticos de compartir y confiar, en los que priman los aspectos técnicos y humanos sobre los informáticos y documentales. Proporciona herramientas y soporte a los profesionales de la información juvenil para aunar sus esfuerzos, conocimientos y tiempos. Actualmente no existe una cultura de compartir recursos entre los centros de información juvenil, ya que cada uno trabaja de forma aislada repitiendo los aciertos y fracasos del resto de los centros. La novedad que se aporta es la posibilidad de compartir la misma información entre los distintos centros de información juvenil. Es una variación española de las tecnologías Content Manager System (CMS), que permite mantener un mismo sistema contenedor de datos desde distintos lugares.

Esta concepción de Internet como un sistema transaccional basado en la confianza y en el compartir libre y voluntario de esfuerzos y conocimientos es la base de los movimientos *antiglobalización*, la comunidad *hacker* y los movimientos de “Mundo Libre” — Software libre, Hardware Libre, Vuelo Libre, Música Libre, Literatura Libre... —.

Está en estudio el factor de las diferentes lenguas que son utilizadas por los usuarios del servicio, ya que al ser un proyecto de buenas prácticas se intenta dar

acogida a todo el ámbito nacional relacionado con la juventud y tanto a usuarios técnicos como a usuarios finales.

El proyecto no es una página web, ni un portal; es una base de datos accesible desde Internet, que pone a disposición de todos los centros que quieran participar unas tablas o registros de información, para poder depositar en ellos la información captada por cada centro, de tal forma que esta pueda ser extraída, explotada, maquetada o investigada, por cualquiera de los centros que estén registrados. Es un proyecto vivo que todavía está en la fase de puesta en marcha y en el que todavía quedan muchas cosas por cambiar y mejorar.

## **7. Referencias**

- Infojoven.com. Zaragoza: Gobierno de Aragón, [s. f.]. URL: <<http://www.infojoven.com>>. Consulta: 24 noviembre 2003.
- Colás, Ana María; Juste, José; Monzó, M. José. Manual de Internet para centros de información juvenil. Madrid: Instituto de la Juventud, 1998.
- Agencia Europea de Información y Asesoramiento para los Jóvenes (ERYICA). Carta europea de información para jóvenes. URL: <<http://www.eryica.org/>>. Consulta 24 noviembre 2003.
- Groff, James; Weinberg, Paul. Guía LAN Times de SQL. Madrid: McGrawHill, 1998.