

Temario de Ayudante de Biblioteca de la Administración General del Estado

Este temario ha sido elaborado por un opositor, para presentarse al proceso selectivo de Ayudante de Bibliotecas de la Administración General del Estado en la [convocatoria de 2021](#).

Incluye todos los temas, de legislación y específicos de bibliotecas, del programa correspondiente a la convocatoria de la Administración General del Estado para cubrir plazas de Ayudante de Bibliotecas en el Ministerios de Cultura y Deporte, Ministerio de Defensa, Ministerio de Asuntos Exteriores, Unión Europea y Cooperación y Ministerio de la Presidencia, Relaciones con las Cortes y Memoria Democrática. «BOE» núm. 149, de 23 de junio de 2021.

Temario completo disponible en:

<https://www.bibliopos.es/>



Temario de Ayudante de Biblioteca de la Administración General del Estado, cedido por su autor a [Bibliopos.es](https://www.bibliopos.es) para su publicación bajo licencia [Creative Commons Reconocimiento-NoComercial-SinObraDerivada 4.0 Internacional License](#).

Bajo esta licencia puedes utilizar libremente el temario para uso personal y compartirlo siempre que [cites la fuente](#) y proporciones un enlace a la [licencia](#). No puedes hacer uso comercial del documento.

B08 Protocolos y lenguajes de búsqueda e intercambio de la información

Z3950, SRU/SRW, OpenUrl, OAI-PMH.

Protocolos y lenguajes de búsqueda e intercambio de la información

Con el surgimiento y desarrollo de las tecnologías de la información y las comunicaciones, se ha logrado almacenar, compartir y gestionar grandes colecciones de datos en formato digital. Estos avances conllevaron al surgimiento de las bibliotecas digitales, cuyo objetivo es el acceso universal a la información, sin límites de tiempo ni espacio, asegurando la disponibilidad, recuperación y autenticidad de los documentos. Para ello, dado el enorme volumen documental producido en la Sociedad del Conocimiento, resulta imprescindible garantizar el éxito de las labores de búsqueda y recuperación de recursos.

Para recuperar la información almacenada, será necesario la traducción de un enunciado en un lenguaje natural a un lenguaje determinado accesible para el sistema. El lenguaje en cuestión se denomina *lenguaje de búsqueda* (también denominados *lenguaje de interrogación* o *de consulta*), y el resultado que se obtiene recibe el nombre de *ecuación de búsqueda*. Los **lenguajes de búsqueda** son la herramienta mediante la cual un usuario expresa su necesidad de información, bien sea en un motor de búsqueda o una base de datos, utilizando filtros o condiciones, para delimitar y precisar los resultados. En otras palabras es la estrategia de búsqueda que un usuario plantea para satisfacer sus necesidades informativas. Son un conjunto de órdenes, operadores y estructuras que, organizadas según unas normas lógicas, permite buscar registros pertinentes que contienen uno o más términos simples o compuestos en la estructura de almacenamiento de un sistema de recuperación de la información. El objetivo principal de todo el proceso es recuperar aquellos documentos que se ajusten de la manera más adecuada a la necesidad de información originaria, es decir, realizar una recuperación de información en la que, en el mejor de los casos, el *ruido* (conjunto de documentos que son recuperados a partir de la ecuación de búsqueda pero que no se adecúan a la necesidad de información originaria) y el *silencio* (conjunto de documentos que no son recuperados a partir de la ecuación de búsqueda pero que sí se adecúan a la necesidad de información originaria) sean mínimos.

Los **operadores** son los encargados de estructurar las ecuaciones (es decir, de formular la pregunta), lo que permite enfocar la búsqueda vinculando términos de búsqueda y definiendo la relación entre ellos. Dentro de los operadores se encuentran los siguientes:

- Operadores booleanos o lógicos. Los operadores booleanos (AND, NOT, OR, XOR) localizan registros que contienen los términos coincidentes en uno o en todos los campos especificados. Las relaciones entre los términos pueden representarse por álgebra de conjuntos o de Boole: suma o unión (or); intersección (and); exclusión (no); unión sin intersección (xor exclusivo).
- Operadores de posición o posicionales. Los operadores de posición (SAME, WITH, NEAR, ADJ, empleo de comillas) localizan registros en los que los términos están próximos, dentro de un mismo registro bibliográfico. Se pueden utilizar para conectar palabras o frases dentro de un campo de búsqueda, pero no entre campos de búsqueda.
- Operadores relacionales o de comparación: Especifican el rango de una búsqueda. Suelen utilizarse en el caso de los campos numéricos.
- Máscaras o Truncamientos: Es un símbolo (asterisco * o interrogante ?) que representa uno o varios caracteres con valor comodín. Se pueden recuperar cadenas de caracteres cuando se desconoce la secuencia ortográfica o algún tipo de variable morfológica. Pueden sustituir un

grupo de caracteres indeterminados al principio, en el medio o al final de una cadena específica, incluso palabras completas.

En principio, cada instrumento de almacenamiento y recuperación de información emplea sus propios sistemas de búsqueda y recuperación de información, caracterizados por unos determinados lenguajes de interrogación así como por unos procedimientos y métodos específicos para su manejo. El problema radica en que los usuarios necesitarían conocer el funcionamiento de cada sistema para poder tener éxito en su tarea de rastreo de la información.

Para evitarlo, hace años que se realizan importantes esfuerzos por desarrollar normas, estándares y protocolos capaces de posibilitar la **interoperabilidad** (capacidad de los sistemas informáticos de interactuar a través del intercambio de información y servicios para lograr un objetivo: el entendimiento entre aplicaciones diversas de búsqueda y localización de información) entre equipos heterogéneos y, con ello, la transparencia en el intercambio de información bibliográfica, tanto a nivel nacional e internacional. En el ámbito de las bibliotecas, la normalización juega un papel crucial, ya que permite simplificar y racionalizar los métodos de trabajo. Lo que hace la normalización es, pues, facilitar las operaciones documentales y agilizar los intercambios de la información, mediante la disposición de las normas correspondientes. Un **protocolo** es un conjunto de normas que regulan el intercambio de información entre ordenadores, normalmente de forma remota. Aunque los ordenadores presenten plataformas informáticas o sistemas operativos diferentes, el protocolo facilita la comunicación entre ellos en un lenguaje común y facilita la interconexión entre los usuarios y las bases de datos necesarias, a partir de una interfaz común de fácil manejo.

Hasta ahora, el mayor obstáculo para las iniciativas cooperativas ha sido, sin duda, la elección de diferentes sistemas de automatización por parte de las bibliotecas, lo que está intentando subsanarse a través de la elaboración de estándares y protocolos unificadores. Las normas, que deben ajustarse a los principios de uniformidad, unificación e identificación, suelen estar elaboradas por organismos internacionales, nacionales, regionales o locales, aunque también por asociaciones particulares e, incluso, empresas.

Z3950

Los ordenadores y las redes de telecomunicaciones ofrecen la posibilidad de compartir datos en formato electrónico. Por lo tanto, se propuso el reto de desarrollar un mecanismo que permitiese estandarizar la comunicación entre sistemas informáticos existentes. En 1988, la NISO (National Information Standards Organization), asociación norteamericana sin fines de lucro acreditada por la ANSI (American National Standards Institute), aprobó la norma **ANSI/NISO Z39.50** (conocida como Z39.50-1988 o la versión 1; en 1992 publicó la versión 2 y en 1995 la versión 3), cuya última versión data de 2003. En 1990 se establecieron dos importantes grupos que garantizan el desarrollo controlado y la continua evolución del estándar: un grupo de implementadores ZIG (Z39.50 Implementors Group) y una agencia para el soporte y mantenimiento del estándar (Z39.50 Maintenance Agency de la Library of Congress). Se adoptó como norma internacional bajo la denominación **ISO 23950:1998 - Information retrieval (Z39.50)**, adquiriendo una dimensión internacional. Es un estándar internacional para la recuperación de la información basado en la estructura cliente/servidor (es decir, existen clientes que hacen requerimientos hacia uno o más servidores, éstos reciben los requerimientos, los procesan y envían al cliente el resultado de dicho procesamiento), que facilita la interconexión entre sistemas informáticos.

Este protocolo hace posible la comunicación entre sistemas que utilizan diferente hardware y software; permite la realización de búsquedas simultáneas a múltiples bases de datos, sin tener que conocer para ello las sintaxis de búsqueda que utilizan dichos sistemas, utilizando una única interfaz de usuario; así como recuperar la información, ordenarla y exportar los registros bibliográficos, mediante la normalización de los procedimientos y las características de búsqueda y

recuperación de información. Normaliza el procedimiento para que dos ordenadores se comuniquen cuando se desea recuperar información. Al quedar establecidas las especificaciones técnicas necesarias para que dicha comunicación pueda producirse, ésta resulta posible aunque los motores de búsqueda o bases de datos sean diferentes, sin necesidad de conocer sus estructuras. El protocolo Z39.50 provee entonces un estándar para el intercambio de mensajes entre el sistema origen y objetivo, la estructura y la semántica de la ecuación de búsqueda, la secuencia del intercambio de mensajes y el mecanismo para devolver los resultados.

Los tres servicios o facilidades principales (de un total de 11) que brinda el protocolo son: la inicialización (donde se establecen los valores iniciales para la ejecución de un programa), la búsqueda (permite realizar búsquedas simples o complejas con la misma herramienta a múltiples bases de datos, agilizando la recuperación de información) y la recuperación de la información (una vez realizada la búsqueda, el cliente solicita al servidor los registros que quiere visualizar, que dependiendo del número solicitado, podrán aparecer segmentados en conjuntos de registro).

Su aplicación en el mundo de las bibliotecas y de los centros de documentación permite el desarrollo de sistemas bibliotecarios interrelacionados, así como la consulta de recursos distribuidos en distintas bases de datos, desde un mismo punto de acceso. Para algunos se trata de la norma más importante para el mundo de las bibliotecas y de la documentación desde la aparición del formato MARC. Las aplicaciones más destacables, a efectos de tareas bibliotecarias, se han centrado fundamentalmente en los aspectos relativos a los OPACs, la catalogación compartida, la construcción de catálogos colectivos, el préstamo interbibliotecario o la difusión selectiva de la información. En el ámbito bibliotecario son muy útiles los “servicios extendidos” que permiten archivar las estrategias y resultados de las búsquedas, actualizar bases de datos, etc. A los usuarios les permite hacer búsquedas y recuperar información de múltiples bases de datos. En la actualidad, la mayoría de los SIGB incorporan este protocolo en sus módulos.

SRU/SRW

El último proyecto promovido por el grupo de implementadores del protocolo Z39.50 fue **ZING** (Z39.50 International Next Generation). Fue ideado con el propósito de promover un conjunto de iniciativas con el objetivo de hacer más accesibles los contenidos tanto intelectuales como semánticos del protocolo para transformarlo en una herramienta más atractiva para desarrolladores, usuarios, proveedores de información y distribuidores. Sus componentes principales son SRU (Search/Retrieval via URL), CQL (Contextual Query Language) y ZeeRex (Z39.50 Explain, Explained and Re-Engineered in XML, que significa “Z39.50 Explain, explicada y remodelada en XML”, que actualiza la utilidad “Explain” del protocolo Z39.50, que permite obtener información de las bases de datos objetivos de la búsqueda): SRU es el protocolo de búsqueda y recuperación de información en Internet, CQL el lenguaje de interrogación y ZeeRex el servicio de descripción. En conjunto, este grupo se llamó originalmente “Z39.50 Next Generation” y, posteriormente, “Z39.50 International Next Generation”. Estos nombres ya no se utilizan.

SRU fue concebido en su origen como una de las parejas de protocolos **SRU/SRW**. SRU y SRW nacieron como los sucesores del protocolo Z39.50 y se apoyan sobre tecnologías actuales y muy difundidas: HTTP (Hypertext Transfer Protocol o Protocolo de Transferencia de Hipertexto, que es el protocolo usado en cada transacción de la World Wide Web) y XML (eXtensible Markup Language, lenguaje de marcado que proporciona una sintaxis superficial para documentos estructurados). Son unos protocolos de búsqueda y recuperación de información desarrollados y mantenidos por la Library of Congress. Su objetivo es normalizar la forma de consultar un servidor de información remoto disponible en la web. En este sentido, SRU/SRW viene a ser una simplificación de la norma Z39.50, adaptada a las características técnicas de la web, basada en el intercambio de peticiones y respuestas XML. En comparación a Z39.50 puro, una implementación basada en SRU/SRW aporta una claridad y sencillez mucho mayor.

SRU (Search/Retrieve via URL) es un protocolo estándar de búsqueda y recuperación de

consultas basado en XML que utiliza el lenguaje de consulta contextual CQL. La versión 1.1 fue lanzada en 2004 y la versión 1.2 en 2007; en 2013 aparece SRU 2.0, un estándar OASIS (Organization for the Advancement of Structured Information Standards, Organización para el Avance de Estándares de Información Estructurada, en castellano). Se caracteriza por enviar tanto las peticiones como las respuestas de búsqueda (y cualquier otra indicación) mediante parámetros dentro de una URL (del inglés Uniform Resource Locator, Localizador Uniforme de Recursos, dirección web) del servidor remoto. Esto es, todos los comandos necesarios para que el servidor entienda una petición y lleve a cabo las acciones pertinentes, se envían dentro de la URL misma de la petición. Consta de dos partes: la URL base (dirección del servidor contra el que se hará la consulta) y la parte de búsqueda (serie de parámetros, separados por el carácter “&”, con la forma “clave=valor”). El servidor recibe el requerimiento, ejecuta la consulta especificada y genera una respuesta con formato XML, la cual es enviada al cliente para que la interprete.

La iniciativa **SRW** (Search/Retrieve Web Service, servicios web de búsqueda y recuperación) es parte de un esfuerzo colaborativo internacional para desarrollar una interfaz estándar de búsqueda de texto basada en la web. La idea es muy similar a la de SRU (en la actualidad, SRW ya no se presenta como un protocolo separado, sino como una variación de SRU). La diferencia radica en que los mensajes entre el cliente y el servidor no son enviados a través de una URL, sino que los mensajes se dirigen del cliente al servidor utilizando el lenguaje de marcado XML sobre el protocolo HTTP, por medio de SOAP (Simple Object Access Protocol, creado y es mantenido por el World Wide Web Consortium, W3C), que especifica cómo encapsular un mensaje dentro de un esquema XML. La norma SRW intenta adherirse a las directrices de la interoperabilidad de servicios web. Los beneficios que se desprenden directamente son un mejor soporte para la extensión, autenticación y características de servicios web. Existen varios proyectos de código abierto que respaldan el estándar SRU/SRW en relación con los principales softwares de repositorio digital, entre ellos DSpace y FEDORA.

OpenUrl

OpenUrl es una sintaxis para transferir metadatos embebidos en enlaces URL de forma ordenada, que suelen conectar con un recurso electrónico. La norma **ANSI/NISO Z39.88-2004 (R2010) The OpenURL Framework for Context-Sensitive Services** (mantenida por OCLC, Online Computer Library Center) define una arquitectura que ayuda a construir direcciones web con información descriptiva sobre un registro (ISSN, volumen, número, DOI u otra información bibliográfica), a fin de que pueda ser localizado el texto completo del mismo, o alguna información relacionada. Su estructura se divide en dos partes: la URL base (dirección del servidor) y la descripción o query (metadatos, separados por el signo &). El **funcionamiento** es el siguiente:

1. Se codifica la referencia bibliográfica de un recurso electrónico dentro de la URL. Cada servidor usa diferentes formatos de OpenURL. Junto al nombre del servidor al que se dirige la petición, se añaden los metadatos sobre el recurso electrónico.
2. El enlace OpenURL no se dirigirá al sitio web del proveedor del recurso, sino a una aplicación informática complementaria llamada **servidor de enlaces** o **resolvedor de enlaces** (*link resolver*, como *360 Link*, implementado junto con el buscador de recursos electrónicos *Summon* de Ex Libris). El servidor de enlaces interpretará los metadatos incluidos en la petición, comprobará la disponibilidad del recurso solicitado en los distintos proveedores de recursos electrónicos a los que el usuario tenga derecho de acceso y devolverá al mismo una página web con el listado de accesos (esta vez, los enlaces apuntarán a los sitios web y bases de datos de cada uno de los proveedores que se hayan identificado). Por tanto, el servidor de enlaces establece así un puente entre una aplicación origen (normalmente un OPAC, aunque podría ser cualquier otra, incluso una base de datos de un tercero) y las bases de datos de los distintos proveedores que ofrecen acceso a los recursos electrónicos o a algún servicio relacionado con ellos (texto completo, resumen,

servicio obtención de documentos, préstamo interbibliotecario, etc.).

OAI-PMH

Acceso abierto (en inglés *open access*) es el acceso inmediato sin requerimiento de registro, suscripción o pago (es decir, sin restricciones) a material digital educativo, académico, científico o de cualquier otro tipo, principalmente artículos de investigación de revistas especializadas con “revisión por pares o iguales” (*peer review*). Es una manera gratuita y abierta de acceder a la literatura científica, aunque también se extiende a otros contenidos digitales que los autores desean hacer libremente accesibles a los usuarios en línea. Así, los pilares sobre los que se apoya este movimiento son, por un lado las revistas científicas que responden a la definición total o parcial de *open access* y, por otro, el depósito, almacenamiento o archivo de trabajos de investigación (pre- o post-print) en repositorios temáticos o institucionales a los que los usuarios, en general, puedan acceder para su búsqueda y recuperación a través de Internet. Con estos repositorios no sólo se consigue facilitar la disponibilidad de los trabajos, sino que aumenta su visibilidad e impacto.

La Open Archives Initiative (OAI, Iniciativa Abierta de Archivos) desarrolla y promueve estándares para lograr la interoperatividad y el acceso a contenidos procedentes de varios proveedores y con diferentes formatos con el fin de facilitar la eficiente difusión de los contenidos. Propuso un protocolo técnico para automatizar la captura y agregación de metadatos de recursos distribuidos en la web, llamado **OAI-PMH** (Open Archive Initiative-Protocol for Metadata Harvesting, Protocolo de Archivos Abiertos para la Recolección de Metadatos), proyecto que dio sus primeros pasos en 1999. Las especificaciones técnicas fueron hechas públicas en 2001 con la publicación de la versión 1.0, elaborándose en 2002 la versión 2 del protocolo. Constituye uno de los pilares técnicos de los llamados archivos abiertos y de los repositorios institucionales. Se podrían señalar tres características fundamentales del protocolo: simplicidad, normalización (basado en estándares ampliamente utilizados en Internet) y recolección (o *harvesting*, recoger los metadatos procedentes de distintos repositorios OAI y almacenarlos en una base de datos).

El Protocolo OAI para la Recolección de Metadatos (OAI-PMH) es una herramienta de interoperabilidad que posibilita el intercambio de metadatos sobre cualquier material almacenado en soporte electrónico. Esta transferencia puede realizarse desde puntos proveedores de servicio a través de búsquedas que abarquen la información recopilada en distintos repositorios de archivos asociados (proveedores de datos). OAI-PMH define un mecanismo para la recolección de registros que contienen los metadatos de los repositorios mediante un proceso automático que se ejecuta periódicamente.

Básicamente, OAI-PMH utiliza transacciones HTTP para emitir preguntas y obtener respuestas entre un servidor o archivo y un cliente o servicio recolector de metadatos. El segundo puede pedir al primero que le envíe metadatos según determinados criterios. En respuesta el primero devuelve un conjunto de registros en formato XML, incluyendo identificadores (URL, por ejemplo) de los objetos descritos en cada registro. Esta estructura de interoperabilidad corresponde al enfoque de *harvesting* (recopilación), en la que los datos y metadatos son transferidos desde la fuente remota al destino en el cual se realizarán los servicios de búsqueda. El *harvesting* es opuesto como estructura de interoperabilidad a la usada por el protocolo Z39.50, cuyo enfoque corresponde al de búsqueda distribuida, es decir, la posibilidad de buscar simultáneamente en múltiples fuentes de datos remotas independientemente de su localización.

El protocolo soporta múltiples formatos para expresar los metadatos, no obstante requiere que todos los servidores ofrezcan los registros utilizando Dublin Core codificado en XML. Además, cada servidor es libre de ofrecer los registros en otros formatos adicionales (MARC, por ejemplo).