

Tutorial de Pubmed

Búsqueda en Medline por e-mail

Diferentes accesos al Medline:

Existen diversas formas de acceder a la base de referencias bibliográficas Medline:

**CD-ROM's*, la "National Library of Medicine" de los Estados Unidos (NLM) ofrece gratuitamente el contenido de la base y licencia a empresas comerciales, el desarrollo del software que se encarga de realizar las búsquedas, con lo cual dichas empresas pueden vender dicho software y no la base en sí misma. Este fue uno de los métodos que posibilitó la difusión de las consultas a la base en las bibliotecas médicas. Posee la desventaja del requerimiento de hardware específico para poner varios años en línea para las consultas y la actualización periódica no es mayor a una vez por mes. Es costosa la suscripción, pero es una solución en ambientes hospitalarios y redes académicas.

**Consultas a expertos* (bibliotecarios, documentalistas, etc.), que poseen la base y generalmente cobran por búsqueda realizada. Tiene la ventaja de que los encargados de realizar las búsquedas, son en general personal calificado y entrenado en búsquedas bibliográficas, con lo cual, si se formula correctamente la petición, los resultados son más representativos.

**Consulta directa* a proveedores del servicio *vía Web*, tiene la ventaja de realizar la búsqueda el propio interesado, la desventaja de la lentitud, según el acceso a internet que se posea, y la falta de destreza o conocimientos para realizar dicha búsqueda. Otro de los problemas es que no todos los integrantes del equipo de salud poseen acceso a internet.

**Consulta diferida por correo electrónico*: su ventaja principal es la de tener menor costo y en que no es necesario tener acceso a "full" internet. Se realiza desde la cuenta de e-mail. Es sin lugar a dudas la primera de las herramientas de internet que llega a lugares remotos. La respuesta es más rápida que en el entorno Web para la realización de las consultas. Dichas consultas se adaptan a la necesidad de cada usuario.

En la presente entrega desarrollaremos un tutorial sobre como realizar las búsquedas de Medline *vía* correo electrónico.

QUERY E-mail server:

Para la realización de las búsquedas en la base de referencias bibliográficas Medline, la NLM ofrece el acceso *vía* Web o *vía* correo electrónico.

El "*QUERY E-mail server*", es un servidor de correo electrónico que permite a los usuarios realizar consultas *vía* e-mail en una serie de bases mantenidas por el "National Center for Biotechnology Information" (NCBI) dependiente de la NLM.

Dicho servidor utiliza el "*Entrez Query Engine*" para obtener los datos. Dicho motor de búsqueda es el mismo que utiliza el PubMed al realizar las consultas al Medline *vía* Web. Las bases a las cuales consulta el mencionado motor de búsqueda son:

- 1) Nucleotide Sequences: GenBank, EMBL, DDBJ, dbEST, dbSTS, dbGSS
- 2) Protein Sequences: GenPept (traducción codificada de regiones de ADN), PIR, SWISSPROT, PRF, PDB.
- 3) 3-D Structures: Molecular Modeling Database (MMDB) database, derived from Brookhaven National Laboratory's Protein Data Bank (PDB)
- 4) **PubMed MEDLINE**: citas bibliográficas de la base completa MEDLINE y PreMedline.

Para realizar las consultas al Medline *vía* e-mail en primer lugar tenemos que tener claro el siguiente concepto:

Interactuaremos con un robot, o mejor dicho, con una computadora que posee un software capaz de comprender determinados comandos que irán escritos en el cuerpo del mensaje de correo electrónico, el cual automáticamente responderá a lo requerido, SIN intervención de humanos en el proceso.

Este concepto es de vital importancia, ya que cualquier error de tipeo o de mala sintaxis del pedido, tornará infructuosa nuestra requisitoria.

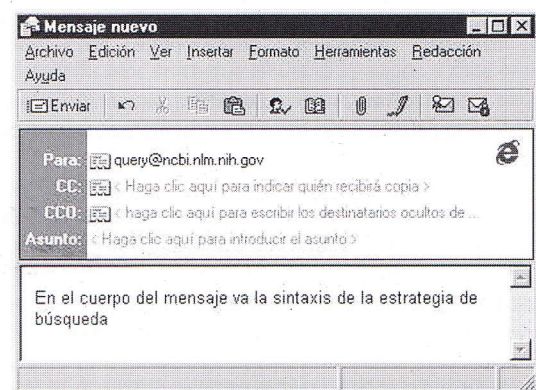
Estructura del mensaje:

En primer término el usuario interesado en realizar la consulta al Medline *vía* correo electrónico debe ingresar al programa gestor del servicio de e-mail, luego debe preparar un mensaje dirigido al servidor de la NLM (el QUERY E-mail server). La **dirección de correo electrónico** de dicho server (el robot con el cual interactuaremos) es la siguiente:

query@ncbi.nlm.nih.gov

Dicha dirección debe ser ingresada en el sector de "Para" o "To" en la cabecera del mensaje; en "Asunto" (en algunos gestores de e-mail se denomina "Subject" o "Tema") NO se debe escribir nada, se debe dejar en blanco dicho sector de la cabecera del mensaje de correo electrónico, al igual que los sectores predeterminados para las copias a terceros de dicho mensaje.

En el **cuerpo del mensaje** (como lo muestra la imagen) se debe escribir la sintaxis de la estrategia de búsqueda. Dicha sintaxis equivale a los comandos que el servidor reconocerá.



En primer término existen dos tipos de búsqueda:

- 1) Búsqueda por identificador único (UID)
- 2) Búsqueda por términos textuales (TERM)

La primera es de utilidad para buscar esencialmente en las otras bases en las cuales el "QUERY E-mail server" puede realizar la búsqueda, aunque también se puede por este tipo de búsqueda hacerlo en la base del Medline.

Con el segundo tipo, el TERM (por términos textuales) es con el cual se realizan las búsquedas en la base del Medline.

Formato de la sintaxis:

El formato de la sintaxis de la estrategia de búsqueda consta de cuatro líneas:

Línea 1: especifica el dominio (**DB**) en donde se realizará la búsqueda, es decir en que base de datos dicha consulta será realizada.

m	MEDLINE
n	nucleotide
p	protein
t	3D structure
s	nucleotide or protein

Siempre en los ejemplos utilizaremos la base del Medline para realizar las búsquedas, por lo cual en la primera línea siempre pondremos: **DB m**

Línea 2: especifica el tipo de búsqueda que se realizará (UID o TERM), seguido de la cadena de búsqueda. Esta línea es la fundamental para obtener resultados exitosos.

En principio (salvo que se indique lo contrario) siempre utilizaremos el tipo de búsqueda por términos, por lo cual la segunda línea siempre comenzará con: **TERM**

A continuación del tipo de búsqueda, se especifican los **términos** (siempre en inglés) y seguido los "especificadores de campos de búsqueda", recordemos que cada cita del Medline posee campos que informan sobre algún aspecto en particular de dicho artículo, éstos campos poseen un nombre específico, más adelante los describiremos. Si **NO** se ingresan especificadores de "campos de búsqueda", la consulta, por defecto, se realiza en todos.

Si se ingresan más de un término de búsqueda se pueden usar **operadores booleanos** para relacionarlos. Si los términos ingresados tienen un espacio separándolos y el motor de búsqueda no los puede interpretar como una frase, automáticamente une los mismos con el operador lógico AND.

Los términos pueden ser truncados. La truncación es el proceso por el cual se encuentran todos los términos que comienzan con una raíz de texto previamente fijado. Poniendo un asterisco (*) al final de la raíz de una palabra (p.e. "bacter*") se recuperarían todos los artículos que contengan palabras que comiencen con dicha raíz (p.e. "bacteria", "bacterium", "bacteriophage", etc.).

Con las frases esto no se realiza, ya que si existe un espacio al finalizar la palabra no es tomada en cuenta para la truncación (p.e. "infect*" no recuperaría "infection control").

Los campos numéricos tales como fecha de publicación pueden ser buscados por un rango de valores.

Línea 3: corresponde a la opción **DOPT**, el cual informa las opciones del formato de presentación de las citas recuperadas.

r	Standard Report Format: (Cita, título, abstract, términos de indexación).
b	Abstract Format: (Cita, título, abstract solamente).
L	MEDLINE Format: (Dos letras indicadoras de cada campo, con la posibilidad de ser importado en bases de referencia bibliográfica).
a	ASN.1 Format: (formato usado para desarrolladores de software).
d	Entrez document summary format: (Título solamente) es el formato de presentación por defecto.
m	Sirve para recuperar artículos relacionados a una cita en particular

Línea 4: contiene **parámetros opcionales** de búsqueda tales como **DISPMAX**, para especificar el número máximo de documentos a ser recuperados, **HTML** para indicar dicho formato, o **PATH** para especificar una dirección de correo electrónico a la cual los resultados deberían ser enviados.

Especificadores de campo de búsqueda:

Los especificadores de campos de búsqueda siempre deben estar entre corchetes [], y pueden o no estar en mayúsculas. Si no se inclu-

ye ningún especificador de campo, por defecto se buscan en todos ellos. Puede o no haber un espacio entre el término y el especificador de campo, pero debe siempre existir un espacio a ambos lados de los operadores booleanos.

Los especificadores de campo de búsqueda son (en orden alfabético):

Campo:	Descripción:	Sintaxis del término:
[AD] ó [AFFL]	Affiliation	Dirección institucional

Comentarios:

Contiene la afiliación institucional y la dirección del primer autor (algunas veces la de otros autores) al momento en que el artículo fue publicado. Este campo se encuentra disponible desde 1988, las citas previas a esta fecha no lo contienen.

Campo:	Descripción:	Sintaxis del término:
[ALL]	All Fields	Cualquier término

Comentarios:

Cubre todos los campos de búsqueda de la base de datos del Medline (es la opción por defecto). Por ejemplo: pneumonia [ALL].

Campo:	Descripción:	Sintaxis del término:
[AU] ó [AUTH]:	Author Name	Apellido y dos iniciales sin punto

Comentarios:

Contiene la lista de los autores de cada artículo, se consignan hasta 10 autores, si son más de diez, el décimo es seguido de la expresión "et al". Generalmente se consigna el primer autor o autor principal solamente. Por ejemplo: Cohen MR [AU].

Campo:	Descripción:	Sintaxis del término:
[EDAT]	Entrez Date	AAAA/MM/DD, AAAA/MM o AAAA

Comentarios:

Fecha de ingreso a la base Entrez, ya que previa a la asignación de los términos MeSH, el artículo forma parte de una base previa denominada PreMEDLINE. La base Entrez es en donde se encuentran las bases donde se realizan las búsquedas. Para ingresar un rango de fechas se debe dividir las mismas con dos puntos (:). Por ejemplo: 1997:1998 [EDAT], para el rango comprendido entre esos años.

Campo:	Descripción:	Sintaxis del término:
[IP] ó [ISSUE]	Issue	Número del Journal

Comentarios:

El número del Journal en el cual el artículo fue publicado.

Campo:	Descripción:	Sintaxis del término:
[TA] ó [JOUR]	Journal Title	Nombre completo, abreviatura del Medline o ISSN

Comentarios:

Este campo representa el nombre del Journal donde fue publicado el artículo. Estos están almacenados en su forma abreviada según lo define el Medline, pero pueden ser ingresados en la extensión completa, abreviada o el número de ISSN. Por ejemplo:

- Completos: Journal of Nutrition [TA]
- Abreviados: j nutr [TA]
- Números de ISSN: 0022-3166 [JOUR]

Campo:	Descripción:	Sintaxis del término:
[LA]	Lenguaje	Lenguaje o primeras tres letras del mismo

Comentarios:

Es el lenguaje en el cual el artículo fue publicado. Cabe destacar que muchos de los artículos que no fueron escritos originalmente en inglés, los "non-English", poseen en la base de datos el abstract en inglés. Por ejemplo: french [LA], o con tres letras para el idioma chino: chi [LA].

Campo:	Descripción:	Sintaxis del término:
[MAJR]	MeSH Major Topic	Término MeSH principal

Comentarios:

Incluye todos los términos MeSH que fueron asignados como los más importantes para cada artículo por los indicadores del MEDLINE de la National Library of Medicine.

Campo:	Descripción:	Sintaxis del término:
[MESH] ó [MS]	MeSH Terms	Término MeSH

Comentarios:

Incluye todos los términos del "Medical Subject Headings" (MeSH), un diccionario de términos controlados utilizado para indicar el MEDLINE. Cada cita del MEDLINE tiene asignado un grupo de términos de este diccionario, los cuales se refieren al tema que versa dicho artículo.

Campo:	Descripción:	Sintaxis del término:
[MDAT]	Modification Date	AAAA/MM/DD, AAAA/MM o AAAA

Comentarios:

Fecha de la última modificación del artículo en la base Medline, desde la fecha de ingreso al sistema, puede ser modificado. Para ingresar un rango de fechas se debe dividir las mismas con dos puntos (:). Por ejemplo: 1997:1998 [MDAT], para el rango comprendido entre esos años.

Campo:	Descripción:	Sintaxis del término:
[PG] ó [PAGE]	Page Number	Número de página

Comentarios:

Es el número de la primera página donde aparece publicado el artículo en el Journal.

Campo:	Descripción:	Sintaxis del término:
[PDAT] ó [DP]	Publication Date	AAAA/MM/DD, AAAA/MM o AAAA

Comentarios:

Contiene la fecha en la que el artículo fue publicado, aparece en el formato año/mes/día. Un año solo (p.e. 1984 [DP]) recuperaría todos los artículos publicados para ese año, un año y un mes (p.e. 1984/03 [DP]) traería todos los de ese mes. Para ingresar un rango de fechas se debe dividir las mismas con dos puntos (:). Por ejemplo: 1997:1998 [PDAT], para el rango comprendido entre esos años.

Campo:	Descripción:	Sintaxis del término:
[PT] ó [PTYP]	Publication Type	Tipo de publicación

Comentarios:

Se refiere a la forma de presentación de un artículo. Por ejemplo: letters, review articles, clinical trials, randomized controlled trials, retracted publications.

Campo:	Descripción:	Sintaxis del término:
[SH]	Subheadings	Tipo de subencabezamiento con nombre completo o con abreviación de dos letras

Comentarios:

Los subencabezamientos son subtipos de categorización de los encabezamientos o términos MeSH (términos que mejor definen el contenido de los artículos), con los subheadings se logra describir aún más el tema de las citas a recuperar. Por ejemplo: hipertension [MH] AND drug therapy [SH].

Campo:	Descripción:	Sintaxis del término:
[WORD] ó [TW]	Text Words	Término a buscar solamente

Comentarios:

Incluye todas las palabras en el título y el abstract, más palabras individuales del MeSH y nombres de sustancias químicas. Por ejemplo: hipertension [WORD].

Campo:	Descripción:	Sintaxis del término:
[TITL] ó [TI]	Title Words	Término a buscar solamente

Comentarios:

Incluye todas las palabras en el título solamente. Por ejemplo: hipertension [TITL].

Campo:	Descripción:	Sintaxis del término:
[VI]	Volume	Número del volumen del Journal

Comentarios:

El número del volumen del Journal donde el artículo fue publicado.

Campo:	Descripción:	Sintaxis del término:
[PMID] y [UI]	Números identificadores	Número del identificador

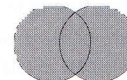
Comentarios:

[UI]: es el identificador único que MEDLINE le asigna a cada registro de la base. [PMID]: es el identificador único que PubMed le asigna a cada registro de la base.

Operadores Booleanos

Los operadores lógicos o Booleanos se basan en la lógica de Boole (matemático inglés del siglo XIX) o teoría de conjuntos. La lógica booleana permite crear y combinar conjuntos de acuerdo a tres criterios básicos: unión, intersección y exclusión, que se traducen en los operadores OR, AND y NOT, respectivamente:

OR: Unión de los elementos de los dos conjuntos de la búsqueda. Es un operador de ampliación. Recupera los recursos que contengan algún elemento de los dos pedidos o los dos.



AND: Es la intersección de los dos conjuntos de búsqueda. Es un operador de reducción. Recupera los recursos que contengan los dos elementos simultáneamente.



NOT: Excluye los elementos de uno de los conjuntos de la búsqueda. Es un operador de reducción. En el resultado solo se recuperarán los recursos que contengan el primero de los elementos y no el segundo.



En la sintaxis de la cadena de estrategia de búsqueda en la línea 2, luego del comando TERM, para combinar términos con los operadores lógicos, SIEMPRE se deben ingresar los mismos en MAYUSCULAS. Recuerde que si no se consigna un identificador de campos de búsqueda, se realiza la misma en todos por defecto. Por ejemplo: hipertension AND therapy al no tener ningún especificador de campo que defina al-

gundo en especial, indica al motor de búsqueda que debe recuperar citas que contengan dichos términos en todos los campos sin distinción. De la misma manera cuando se especifican campos a buscar la sintaxis sería:

término [campo] Booleano término [campo] Booleano...

Las expresiones booleanas son procesadas por el motor de búsqueda de izquierda a derecha. Si se desea, al igual que las ecuaciones matemáticas, se pueden encerrar sentencias entre paréntesis, con lo cual dicha sentencia será procesada como una unidad.

(término [campo] Booleano término) booleano término [campo]...
(hypertension [MH] AND enalapril) NOT Comment [PT]

Formato de presentación:

El comando de la línea 3 le indica al servidor donde se realiza la consulta, el formato en el cual las citas deben ser presentadas.

El formato por defecto en el tipo de búsqueda TERM, es el "d", mostrando solamente el título. Es recomendable usar el formato "r" (Standard Report Format) para la recuperación de citas en particular o cuando se va a recuperar pocas citas. Para la recuperación de muchas citas o el pedido de artículos relacionados, se recomienda usar el "d" ya que de esta manera solamente viene el título y uno puede seleccionar los que le interesan y con el identificador de cita realizar nuevamente la consulta ya en un formato de presentación más ampliado.

Estos identificadores, el PMID o el UI, vienen consignados en el formato por defecto (el "d") con lo cual es recomendable anotar uno de ellos y si se desea ampliar la información con artículos relacionados a una cita en especial se debe combinar la letra de formato base más la letra de artículo relacionado por ejemplo:

DB m
TERM 99040699 [UI]
DOPT dm
DISPMAX 20

con lo cual recuperaremos de la base Medline, los artículos relacionados con la cita identificada con el número 99040699, y que no vengán más de 20 citas en el e-mail de respuesta.

Es posible omitir la línea 3, es decir, dejar por defecto que las citas recuperadas vengán en el formato de solamente título.

Parámetro de recuperación:

Con DISPMAX se indica el número máximo de citas a recuperar; si espera una pequeña cantidad de citas, este comando puede ser omitido, con lo cual, por defecto la cantidad máxima se sitúa en 200. Igualmente tenga en cuenta que la máxima cantidad de líneas que pueden ser contenidas en un mensaje de e-mail, por este servicio, son 100.000 (líneas, no citas), con lo cual dependiendo del formato de presentación (línea 3) vendrán más o menos citas.

Con HTML se le pide al servidor que incluya en el texto marcadores del lenguaje de marcación de hipertexto o HTML (lenguaje para la escritura de páginas Web) con lo cual dicho mensaje puede luego ser visto con un navegador o browser.

Con PATH se indica una dirección de e-mail a la cual usted desea que la respuesta sea enviada. Ej. **PATH dluna@intramed.net.ar**

Algunos Ejemplos Prácticos:

Búsqueda en el Medline del término hipertensión solamente con todas las opciones por defecto:

DB m
TERM hypertension

Analicemos la sintaxis, en la primera línea se especifica la base, en este caso, "m" para Medline, en la línea 2 el tipo de búsqueda, TERM por términos, con el término "hypertension" (siempre deben ser ingresados en inglés), al no incorporar ningún especificador de campo, se buscará en todos

los campos por defecto. En la línea 3 (DOPT) si no se ingresa nada, por defecto, se mostrarán los resultados en el formato sólo de títulos. En la línea 4 de la misma manera al no escribir nada por defecto, no se limita la cantidad de citas a recuperar.

Ahora bien, debido a que con la formulación anterior vinieron, por no acotar la búsqueda, 200 citas, refinaremos la misma. En primer lugar indicaremos que la búsqueda se realice en el campo MeSH, con lo cual nuestro término será buscado en dicho campo y no en los otros. Pidiendo que recupere solamente 20 citas, para lo cual agregamos el comando DISPMAX:

DB m
TERM hypertension [MS]
DISPMAX 20

A continuación refinaremos aún más nuestra estrategia y pediremos que solamente nos recupere artículos del tipo "Revisión" y que se nos muestre el resumen o abstract de los artículos:

DB m
TERM hypertension [MS] AND review [PT]
DOPT b
DISPMAX 20

Debido a que sería conveniente refinar aún más la búsqueda con un enfoque terapéutico, podríamos pedir que a dicho término MeSH se le agregue el subheading o subencabezamiento de tratamiento. Debido a lo extenso de la respuesta del mail previo, nuevamente pedimos el formato por omisión y no incorporamos la línea 3 (también podríamos asignar el parámetro a "b" al comando DOPT):

DB m
TERM (hypertension [MS] AND therapy [SH]) AND review [PT]
DISPMAX 20

Con el ejemplo anterior se recuperaron 20 artículos que tienen el término MeSH hipertensión asignado, con el subencabezamiento therapy y pedimos artículos de revisión, en el formato de sólo título. Bien, ahora refinemos el rango de fecha de publicación de dichos artículos a un año:

DB m
TERM ((hypertension [MS] AND therapy [SH]) AND review [PT]) AND 1998 [PDAT]
DISPMAX 20

Entre las respuestas encontradas con la consulta anterior recuperamos el siguiente artículo:

Calcium antagonists and cardiovascular risk in diabetes. Am J Cardiol. 1998 Nov 12;82(9B):42R-44R. Review. PMID: 9822145; UI: 99037939. 9822144 Poulter NR.

Suponiendo que dicho artículo representa muy bien lo que está buscando, sería conveniente pedir los artículos relacionados, el motor de búsqueda compara las palabras del título y del abstract tanto como los términos MeSH de cada cita con un poderoso algoritmo de comparación de palabras. Por medio de éste, selecciona los artículos que coinciden mejor con el evaluado. El mejor resultado es guardado y almacenado en un set precalculado. La opción de recuperar dichos artículos con el comando DOPT m, trae en orden de relevancia y en cronología decreciente los artículos relacionados.

Si usted encontró uno o más artículos que describen exactamente lo que busca, usando esta función incrementará la especificidad de su búsqueda en relación a las vías convencionales. Siguiendo con nuestro ejemplo tendríamos que tomar el identificador de artículo en la base del Medline y pedir los artículos relacionados, en este caso con el formato convencional de sólo título para poder evaluar los que se desean ampliar posteriormente:

DB m
TERM 99037939 [UI]
DOPT dm
DISPMAX 20

Para comparar las estrategias previamente descritas, si lo desea puede hacer lo mismo vía Web en el PubMed y ver si obtiene los mismos resultados.

Dr. Daniel Luna

Clínica Médica. Hospital Italiano de Buenos Aires.