



Jornadas de
**Visibilidad
Web** UNAM 2021

Alternativas para buscadores
universitarios

Mario Alberto Arredondo



CONTEXTO

- Diversidad de los contenidos y su etiquetado en el dominio unam.mx
- Plataformas en las que se encuentran publicados los contenidos (no todos en el dominio institucional).
- Una enorme cantidad de sitios, páginas y repositorios universitarios (más de 12 mil).
- Para los materiales, no en todos los casos se tienen las referencias en los sitios web principales.
- Tener en cuenta la usabilidad de las interfaces de búsqueda.
- Encontrar soluciones de acuerdo a las necesidades, recursos tecnológicos y capital humano.

TIPOS DE BUSCADORES

- Índices o Directorios
- Jerárquicos
- Semánticos
- Metabuscadores
- Buscadores verticales



ÍNDICES O DIRECTORIOS

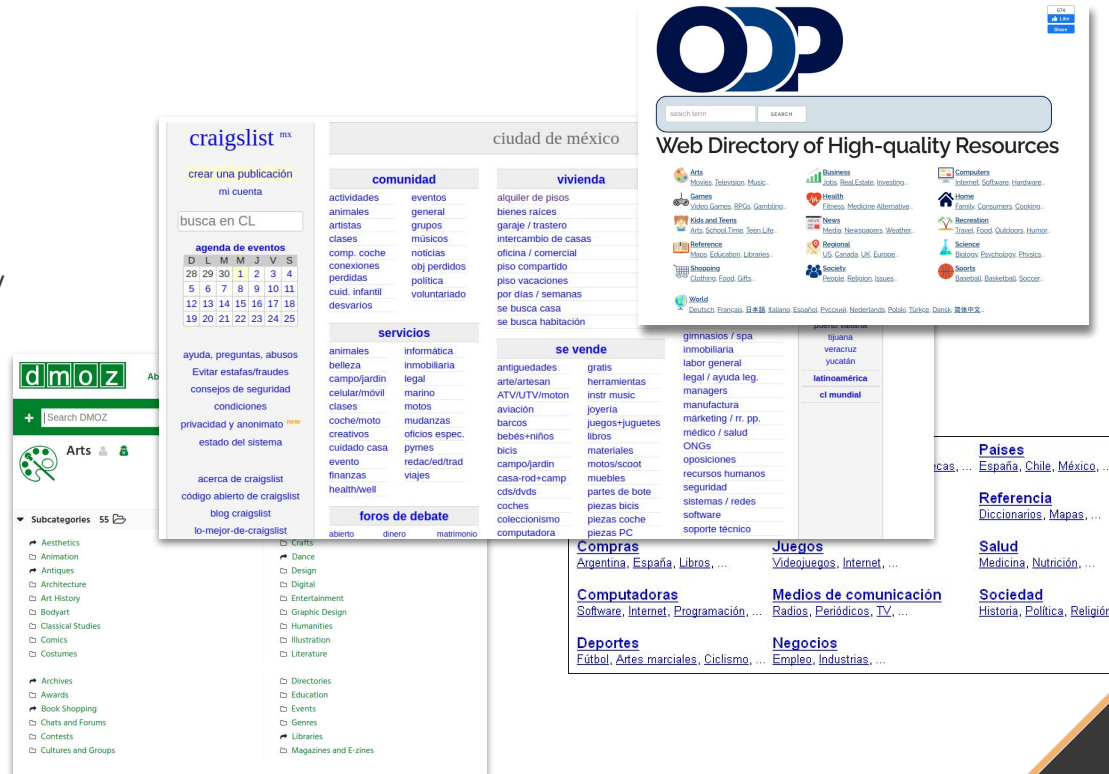
Son directorios de páginas web que regularmente se ordenan por tema. Son simples de usar y se pueden navegar mediante el acceso a categorías y subcategorías.

Generalmente son humanos los que indexan a una base de datos las páginas, según sus contenidos y categorías. El *Open Directory Project* y la *Internet Public Library* son ejemplos de directorios.

EJEMPLOS DE DIRECTORIOS

Otros ejemplos:

- About Us
- DMOZ
- Yahoo Directory
- Craigslist

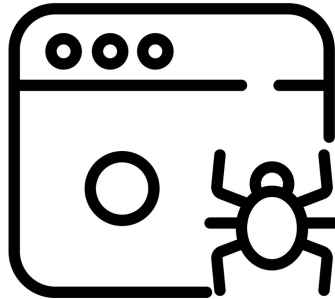


The collage displays three prominent web directories:

- Craigslist:** A classifieds website for Mexico City, featuring sections for community, housing, services, and forums. It includes a search bar and a calendar for events.
- DMOZ (Open Directory Project):** A multi-language web directory with a search bar and a list of subcategories such as Arts, Animation, and Antiques.
- ODP (Open Directory Project):** A web directory of high-quality resources, organized into categories like Arts, Business, Health, and Science, with a search bar and a list of subcategories.

BUSCADORES JERÁRQUICOS

Buscan y almacenan información a través de robots (*crawlers*) que organizan la información de forma jerárquica, de acuerdo al grado de relevancia de los contenidos y al historial de las búsquedas realizadas anteriormente por los usuarios.



EJEMPLOS DE BUSCADORES JERÁRQUICOS

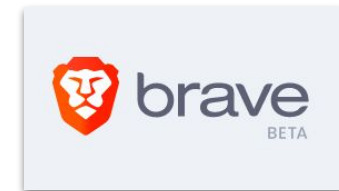
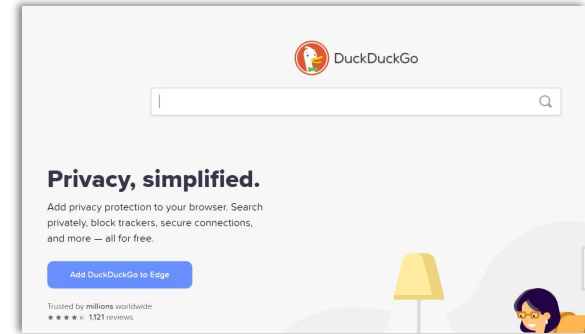
- Google
- Baidu
- AOL
- Ask Jeeves
- Teoma
- Yahoo
- Brave
- Ecosia
- DuckDuckGo
- Bing



The search engine that
plants trees.

138,974,856

Trees planted by Ecosia users



BUSCADORES SEMÁNTICOS

- Extraen la información de los datos etiquetados, cabeceras, RDFa, microformatos.
- Generan ontologías que definen la relación entre las entidades mediante el Lenguaje de Ontología Web (*OWL*).
- Hacen uso de algoritmos que pueden procesar ontologías y datos para crear nuevo conocimiento.
- Lo importante de la web semántica no son las entidades sino cómo se forman y relacionan para dar sentido y poder así obtener resultados adecuados a lo que buscamos

EJEMPLOS DE BUSCADORES SEMÁNTICOS

Open Semantic Search
opensematicsearch

Search, analyze and explore large document collections by Open Source Search Engine, Text Mining, Document analysis and Text Analytics Explorer

598 followers · 14 following · 101

<https://opensematicsearch.org>

Achievements

Block or Report

Overview Repositories 22 Projects

Pinned

- open-semantic-search** (Public)

Open Source research tool to search, browse, analyze and explore document collections by Semantic Search Engine and Open Source Mining & Text Analytics platform (integrates ETL for docum...

● Dockerfile ☆ 479 ▼ 84
- open-semantic-visual-graph-explorer** (Public)

Open Semantic Visual Linked Data Graph Explorer: Open Source tool for user interface (UI) for discovery, exploration and visualization of direct and indirect connections between named entit...

● HTML ☆ 55 ▼ 10
- open-semantic-entity-search-api** (Public)

Open Source REST API for named entity extraction, named entity link, named entity disambiguation, recommendation & reconciliation of entities like persons, organizations and places for (ge)naut...

● Python ☆ 118 ▼ 24

6 contributions in the last year

Calendar: Dec, Jan, Feb, Mar, Apr

WolframAlpha computational intelligence.

Enter what you want to calculate or know about

NATURAL LANGUAGE MATH INPUT EXTENDED KEYBOARD EXAMPLES UPLOAD RANDOM

NEW Use textbook math notation to enter your math. Try it

Answers using Wolfram's breakthrough knowledgebase and AI technology

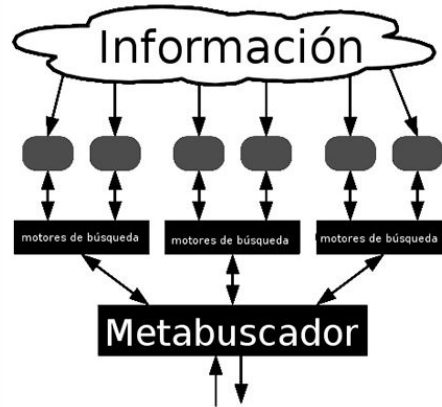
Mathematics > Technology > Society & Culture > Everyday Life >

Step-by-Step Solutions	Units & Measures	People	Personal Health
Elementary Math	Physics	Arts & Media	Personal Finance
x^2-1 Algebra	Chemistry	18 Dates & Times	Surprises
Plotting & Graphics	Engineering	Words & Linguistics	Entertainment
Calculus & Analysis	Computational Sciences	Money & Finance	Household Science
Geometry	Earth Sciences	Food & Nutrition	Household Math

METABUSCADORES

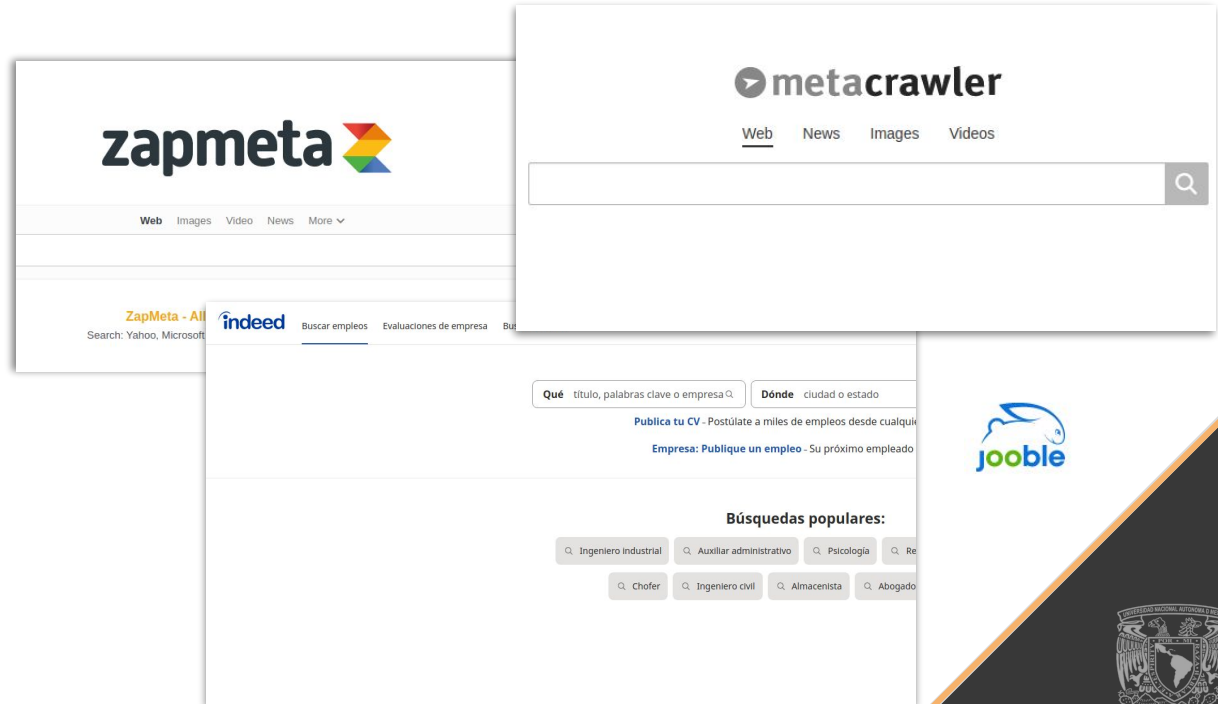
Son buscadores que hacen uso de otros buscadores para ofrecer un resumen de los mejores resultados, evitando mostrar información duplicada.

Permiten ejecutar búsquedas de información en diferentes fuentes electrónicas, de una sola vez y un mismo punto.



EJEMPLOS DE METABUSCADORES

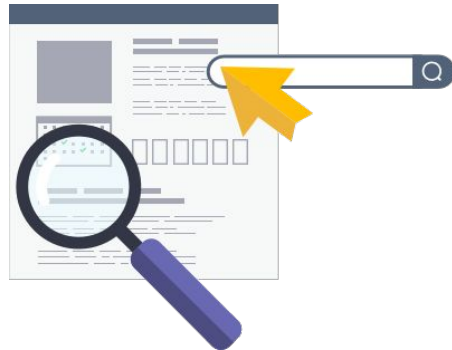
- Jooble
- iBoggie
- Metacrawler
- Indeed
- Zapmeta
- Yippy



BUSCADORES VERTICALES

Pueden considerarse una derivación de los buscadores jerárquicos debido a que, la mayoría, funciona a través de un sistema de arañas o crawlers.

Su mayor diferencia es que se enfocan en una temática específica, es por ello que son mucho más precisos. Teniendo en cuenta que los bots se centran en las páginas que se alinean a un tema determinado, el trabajo de búsqueda resulta mucho más eficiente.

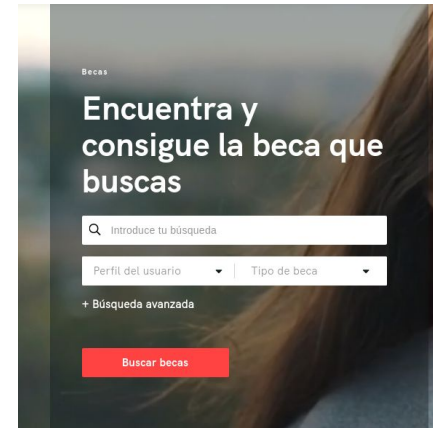


EJEMPLOS DE BUSCADORES VERTICALES

- Google académico
- ARQ
- Becas
- Becas Universia
- BoardReader
- IMDB
- AllMusic

Google Académico

Cualquier idioma Buscar sólo páginas en español



CÓMO SELECCIONAR EL MÁS ADECUADO

Podríamos plantearnos preguntas como:

- ¿Dónde se encuentra almacenada la información?
- ¿Se cuenta con algún tipo de estándar para la clasificación?
- ¿El buscador cuenta con una API para ampliar su funcionalidad?
- ¿Se puede realizar el filtrado según nuestras necesidades particulares?
Por ejemplo: por dominio, tipo de contenido, fecha, región, autor.
- ¿Cuál es el tipo de licencia del buscador?
- ¿Qué plataforma tecnológica necesita para implementarse?
- ¿Cuenta con documentación y soporte adecuados?

CONTENIDO CORRECTAMENTE ETIQUETADO

“ Las metaetiquetas a nivel de página son una forma perfecta con la que los propietarios de sitios web pueden facilitar a los buscadores información sobre sus sitios

```

<!DOCTYPE html>
<html>
  <head>
    <meta charset="utf-8">
    <meta name="Description" CONTENT="Author: A.N. Author, Illustrator:
    <meta name="google-site-verification" content="+nxGUDJ4QpAZ519Bsjd
    <title>Example Books - high-quality used books for children</title>
    <meta name="robots" content="noindex,nofollow">
  </head>
</html>
  
```

<https://developers.google.com/search/docs/advanced/crawling/special-tags?hl=es>

LO QUE EL USUARIO VE

The screenshot shows the homepage of the 'Toda la UNAM en línea' website. At the top, there is a navigation bar with the logo 'Toda la UNAM en línea' on the left, a central menu with items: 'Docencia Digital', 'Cultura', 'Ciencia', 'Repositorio Institucional', and 'Bibliotecas Digitales', and social media icons for Facebook, Twitter, and YouTube on the right. Below the navigation bar is a large banner image of a tree with a search bar overlaid. The search bar contains the text '¿Qué te gustaría explorar en Toda la UNAM en Línea?' and a magnifying glass icon. Below the banner, the page is divided into three columns: 'Destacado', 'Recientes', and 'Materiales más populares'. The 'Destacado' section features a purple banner for 'Facultad de PSICOLOGÍA' with the text 'Difusión', 'Publicaciones Periódicas', and 'Repositorios'. Below this is the title 'Difusión de la Facultad de Psicología' and a short description. The 'Recientes' section lists two items: '41 tropiezos Heteronorma' (CULTURA, 0.0 stars, HTML) and 'La ciencia que somos, Iberoamérica al aire 2021' (CIENCIA, 0.0 stars, HTML). The 'Materiales más populares' section lists three items: '¿Qué tan buena es tu ortografía?' (306180 views, 1.3 stars), 'Exámenes de comprensión de lectura para autoaprendizaje' (71428 views, 4.0 stars), and 'Biblioteca Médica Digital' (69969 views, 0.9 stars). Below these are 'Descarga Cultura UNAM' (45860 views, 10.0 stars) and '100 Unidades didácticas interactivas de las áreas de matemáticas v física'.

Toda la UNAM en línea

Docencia Digital Cultura Ciencia Repositorio Institucional Bibliotecas Digitales

Una entrada a los materiales en línea que la UNAM tiene para ti

¿Qué te gustaría explorar en Toda la UNAM en Línea?

Destacado

Facultad de PSICOLOGÍA

Difusión

Publicaciones Periódicas

Repositorios

Difusión de la Facultad de Psicología

Página que ofrece la consulta de publicaciones y repositorios con los que cuenta la Facultad de Psicología.

Recientes

41 tropiezos Heteronorma

CULTURA ★ 0.0 HTML

Ver ficha técnica

La ciencia que somos, Iberoamérica al aire 2021

CIENCIA ★ 0.0 HTML

Ver ficha técnica

Materiales más populares

¿Qué tan buena es tu ortografía?

306180 ★ 1.3

Exámenes de comprensión de lectura para autoaprendizaje

71428 ★ 4.0

Biblioteca Médica Digital

69969 ★ 0.9

Descarga Cultura UNAM

45860 ★ 10.0

100 Unidades didácticas interactivas de las áreas de matemáticas v física

<https://unamenlinea.unam.mx/>

LO QUE LOS RASTREADORES VEN

```
1 <!DOCTYPE html>
2 <html lang="es">
3 <head>
4   <meta charset="UTF-8">
5   <meta name="viewport" content="width=device-width, initial-scale=1.0">
6   <meta http-equiv="X-UA-Compatible" content="ie=edge">
7   <title> Toda la UNAM en Línea </title>
8   <link rel="icon" href="/images/Favicon.png" type="image/png">
9   <link rel="stylesheet" href="/vendors/bootstrap/bootstrap.min.css">
10  <link rel="stylesheet" href="/vendors/bootstrap/bootstrap.css.map">
11  <link rel="stylesheet" href="https://use.fontawesome.com/releases/v5.10.1/css/all.css">
12  <link rel="stylesheet" href="https://use.fontawesome.com/releases/v5.10.1/css/v4-shims.css">
13  <link href="https://fonts.googleapis.com/css?family=Assistant:300,400,500,600,700" rel="stylesheet">
14  <link rel="stylesheet" href="/vendors/themify-icons/themify-icons.css">
15  <link rel="stylesheet" href="/vendors/linericon/style.css">
16
17  <link rel="stylesheet" href="/vendors/flat-icon/font/flaticon.css">
18
19  <link href="/favicon.ico" type="image/x-icon" rel="icon" /> <link href="/favicon.ico" type="image/x-icon" rel="shortcut icon" />
20  <script async src="https://www.googletagmanager.com/gtag/js?id=UA-11155928-8"></script>
21  <script>
22    window.dataLayer = window.dataLayer || [];
23    function gtag(){dataLayer.push(arguments);}
24    gtag('js', new Date());
25
26    gtag('config', 'UA-11155928-8');
27  </script>
28
29  <!-- Global site tag (gtag.js) - Google Analytics -->
30  <script async src="https://www.googletagmanager.com/gtag/js?id=UA-48070843-1"></script>
31  <script>
32    window.dataLayer = window.dataLayer || [];
33    function gtag(){dataLayer.push(arguments);}
34    gtag('js', new Date());
35
36    gtag('config', 'UA-48070843-1');
37  </script>
38  <script src="/vendors/bootstrap/bootstrap.bundle.min.js"></script>
39  <script src="/vendors/Magnific-Popup/jquery.magnific-popup.min.js"></script>
40
41
42 </head>
```

<https://unamenlinea.unam.mx/>

METAETIQUETAS

- Palabras clave (*keywords*)
- Descripción (*description*)
- Datos estructurados *Json-LD*
- Texto alternativo *ALT*
- Etiqueta canónica (*canonical*)



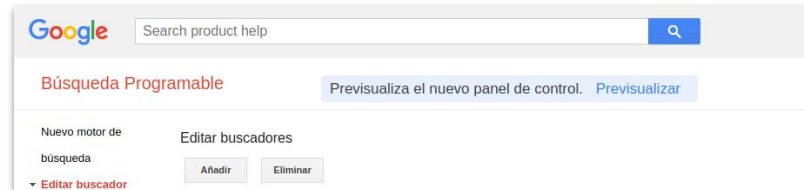
`<meta name="Nombre del elemento" content="Contenido asignado"/>`

ALGUNAS METAETIQUETAS RELEVANTES

- *keywords*: Dan la posibilidad de que los administradores ofrezcan información clave para el motor de búsqueda, sobre el contenido de su sitio.
<meta name="keywords" content="Cursos, Talleres, Diplomados, a Distancia, en línea, presenciales, tecnología, cómputo, educación en línea." />
- *Description*: Se usa para proporcionar una descripción breve de la página.
<meta name="description" content="Cursos y Diplomados en TIC, en las modalidades de educación continua, presenciales y en línea. />
- *Robots*: Controla el comportamiento de rastreo y de indexación de los motores de búsqueda.
<meta name="robots" content="..., ..." />
- *Viewport*: Indica al navegador cómo renderizar páginas en un dispositivo móvil.
<meta name="viewport" content="...">

PROGRAMMABLE SEARCH ENGINE (I)

“ El Motor de búsqueda programable de Google es una plataforma proporcionada por Google que permite a los desarrolladores web presentar información especializada en búsquedas web, refinar y clasificar consultas y crear motores de búsqueda personalizados, basados en la Búsqueda de Google.



<https://programmablesearchengine.google.com/>

PROGRAMMABLE SEARCH ENGINE (II)

Hay dos posibilidades para configurarlo:

- A. Implementarlo para un solo dominio o sitio web.
- B. Implementarlo para que busque en varios sitios web especializados o en un tema en particular.



Si se conoce bien al usuario se puede utilizar la experiencia propia en determinado tema para configurarlo e indicar qué sitios web buscar, priorizar o ignorar.

PROGRAMMABLE SEARCH ENGINE (III)

Con este motor de búsqueda, entre otras cosas, se puede:

- Hacer la búsqueda en una colección específica de sitios o páginas.
- Habilitar la búsqueda de imágenes para un sitio.
- Personalizar la apariencia de la página de resultados.
- Incluir la adición de autocompletado de búsqueda.
- Aprovechar los datos estructurados de un sitio para personalizar los resultados de búsqueda.

CASO PRÁCTICO A. Implementarlo para un solo dominio o sitio web - Toda la UNAM en Línea



Una entrada a los materiales en línea que la UNAM tiene para ti

danza



Alrededor de 47 resultados en Toda la UNAM en Línea



Al son de los clásicos: la danza desechada de un hombre

Artículo que hace referencia a una obra coreográfica realizada en Colombia como los boleros. Dentro del documento el alumno podrá conocer la trama de la puesta en escena y una narración de los cuadros de la obra coreográfica.

Ver ficha técnica



Alineación Corporal

En esta capsula conocerás el concepto de alineación corporal, comprenderás la importancia del mismo, y en el último, se explicará la funcionalidad de la postura.

Ver ficha técnica

Resultados en dominio unam.mx

Aproximadamente 78,000 resultados (0.25 segundos)

Danza | Casa del Lago UNAM

[casadellago.unam.mx](#) > [nuevo](#) > [cursos-y-talleres](#) > [danza](#)



Workout: ballet desde casa · [Danza balinesa](#): la comunidad que baila (Indonesia) · [Danza experimental](#) desde el zapateado tradicional jarocho · [Breaking](#) y freestyle ...

NOSOTROS — Danza UNAM

[Danza UNAM](#) > [nosotros](#)



La Dirección de Danza de la UNAM. Impulsa proyectos de programación artística, actividades académicas y programas recreativos, enfocados a la promoción y ...

2° Diplomado en Danza y Mediación Tecnológica — Danza UNAM

[Danza UNAM](#) > [diplomadodanzaymediaciontecnologica](#)



La Coordinación de Difusión Cultural, a través de la Dirección de Danza UNAM, convoca a creadores e investigadores en disciplinas relacionadas con la escena ...

Danza UNAM

[Danza UNAM](#)



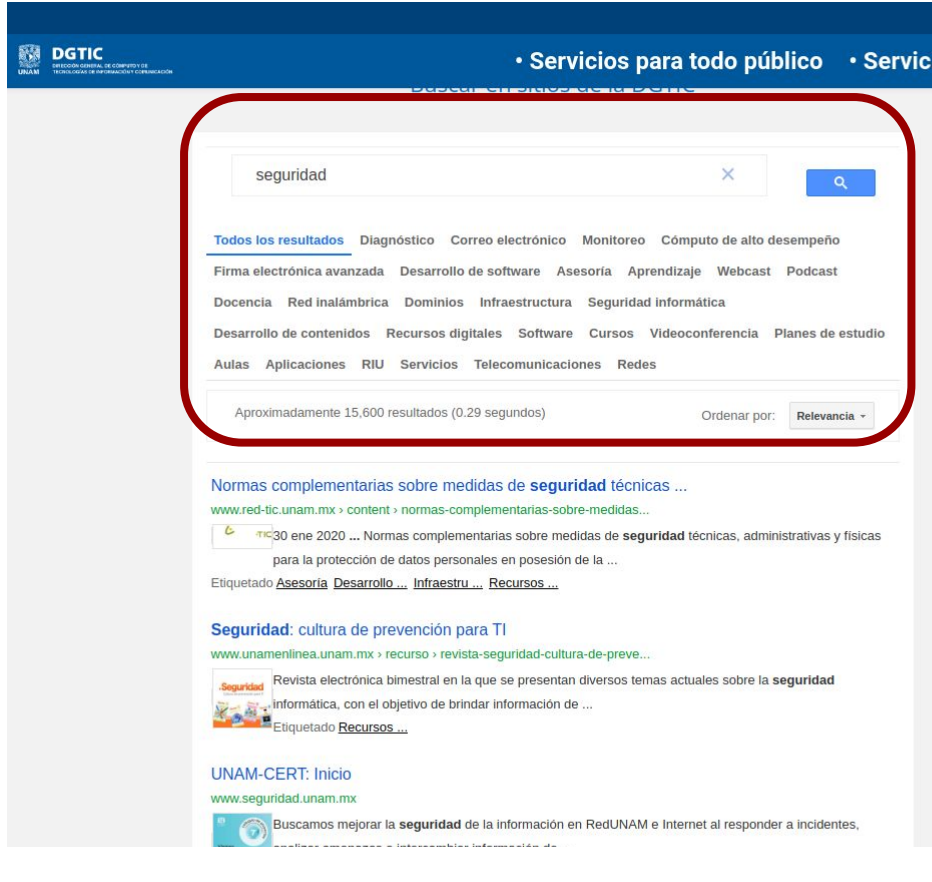
Página oficial de la Dirección de Danza de la UNAM. Impulsa proyectos de programación artística, actividades académicas y programas recreativos, ...

CONVOCATORIA CUERPOS MIGRANTES — Danza UNAM

[Danza UNAM](#) > [convocatoria](#) > [cuerposmigrantes](#)



La Universidad Nacional Autónoma de México a través de la Coordinación de Difusión Cultural y la Dirección de Danza, convoca a la comunidad dancística a ...



DGTIC DIRECCIÓN GENERAL DE COMPUTO Y TIC
 TECNOLOGÍAS DE INFORMACIÓN Y COMUNICACIÓN

Inicio Nosotros

• Servicios para todo público • Servicios

Buscar en los sitios web de DGTIC

seguridad

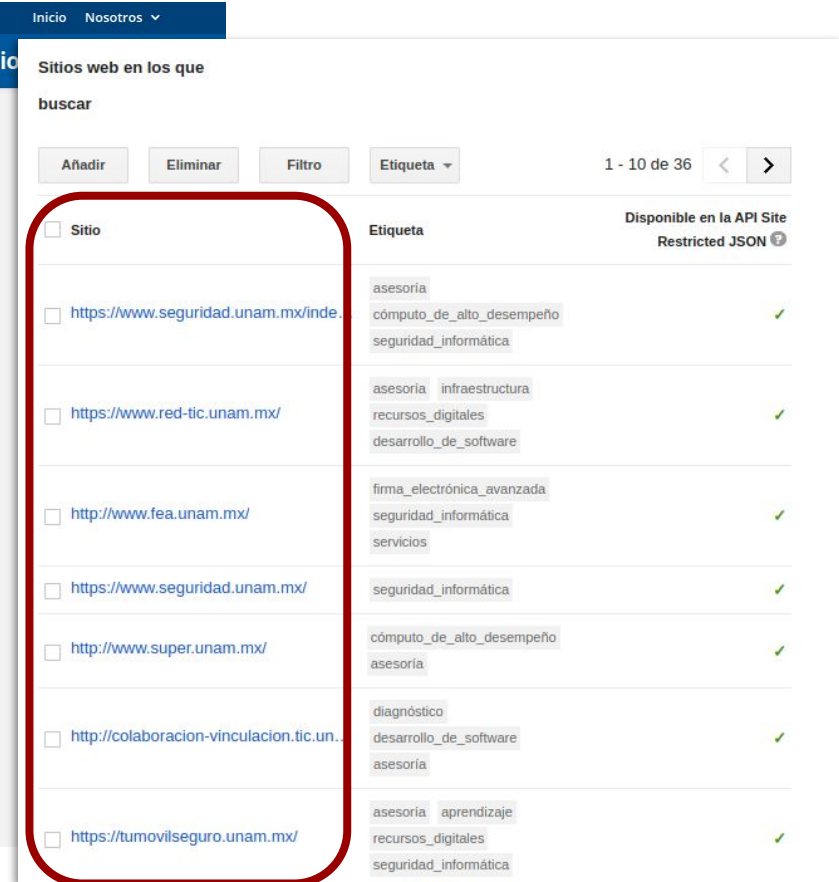
Todos los resultados Diagnóstico Correo electrónico Monitoreo Cómputo de alto desempeño
 Firma electrónica avanzada Desarrollo de software Asesoría Aprendizaje Webcast Podcast
 Docencia Red inalámbrica Dominios Infraestructura Seguridad informática
 Desarrollo de contenidos Recursos digitales Software Cursos Videoconferencia Planes de estudio
 Aulas Aplicaciones RIU Servicios Telecomunicaciones Redes

Aproximadamente 15,600 resultados (0.29 segundos) Ordenar por: Relevancia

Normas complementarias sobre medidas de **seguridad** técnicas ...
[www.red-tic.unam.mx](#) > content > normas-complementarias-sobre-medidas-...
 30 ene 2020 ... Normas complementarias sobre medidas de **seguridad** técnicas, administrativas y físicas para la protección de datos personales en posesión de la ...
 Etiquetado [Asesoría](#) [Desarrollo](#) ... [Infraestructura](#) ... [Recursos](#) ...

Seguridad: cultura de prevención para TI
[www.unamenlinea.unam.mx](#) > recurso > revista-seguridad-cultura-de-preve-...
 Revista electrónica bimestral en la que se presentan diversos temas actuales sobre la **seguridad** informática, con el objetivo de brindar información de ...
 Etiquetado [Recursos](#) ...

UNAM-CERT: Inicio
[www.seguridad.unam.mx](#)
 Buscamos mejorar la **seguridad** de la información en RedUNAM e Internet al responder a incidentes, ...



Sitios web en los que buscar

Añadir Eliminar Filtro Etiqueta 1 - 10 de 36

Sitio	Etiqueta	Disponible en la API Site Restricted JSON
<input type="checkbox"/> https://www.seguridad.unam.mx/index.php	asesoria cómputo_de_alto_desempeño seguridad_informática	✓
<input type="checkbox"/> https://www.red-tic.unam.mx/	asesoria infraestructura recursos_digitales desarrollo_de_software	✓
<input type="checkbox"/> http://www.fea.unam.mx/	firma_electrónica_avanzada seguridad_informática servicios	✓
<input type="checkbox"/> https://www.seguridad.unam.mx/	seguridad_informática	✓
<input type="checkbox"/> http://www.super.unam.mx/	cómputo_de_alto_desempeño asesoria	✓
<input type="checkbox"/> http://colaboracion-vinculacion.tic.unam.mx/	diagnóstico desarrollo_de_software asesoria	✓
<input type="checkbox"/> https://tumovilseguro.unam.mx/	asesoria aprendizaje recursos_digitales seguridad_informática	✓

Buscar
idiomas

Me interesan más los videos

Ciencias Físico - Matemáticas y de las
Ingenierías

Ciencias Biológicas, Químicas y de la
Salud

Ciencias

Circunferencia

Funciones trigonométricas:

[Ley de De Morga

Ver todos los resultados Youtube

Aproximadamente 36,800 resultados (1.35 segundos)

Videos Educativos – CETA

ceta.zaragoza.unam.mx > edu-vids



Videos con enfoque educativo relevante a las carreras de la FES Z y temas complementarios de ... Efectiva herramienta para la enseñanza del Idioma inglés.

Recursos | Colegio de Ciencias y Humanidades

cch.unam.mx > recursos



SSAA VIDEOS. 5 subscribers. Subscribe · Ambiente Virtual de Idiomas. Info. Shopping. Tap to unmute. If playback doesn't begin shortly, try restarting your ...

Aulas Virtuales – CETA

ceta.zaragoza.unam.mx > aulas-virtuales



... los estudiantes solicitan mayor uso de recursos tecnológicos como aulas virtuales o incorporación de recursos como videos, o paquetería de office, ...

Recursos Digitales. Alumno | Colegio de Ciencias y Humanidades

www.cch.unam.mx > recursos > alumno



Ambiente Virtual de IdiomasRefuerza tu aprendizaje en idiomas extranjeros con recursos que ofrece la CUAED (Coordinación de Universidad Abierta y Educación ...

CETED FES Acatlán - YouTube

YouTube > CETEDFESAcatlan > feed



Canal de videos del Centro Tecnológico para la Educación a Distancia de la Facultad de Estudios Superiores Acatlán de la UNAM.

| Biné

bine.iztacala.unam.mx



Recuperado de: https://bine.iztacala.unam.mx/video/ Clausura. Resumen ... Educación y Cultura Libre, Segundo Encuentro en Línea CHAT. [Video].

Recursos Digitales. Profesor | Colegio de Ciencias y Humanidades

Todos los resultados Youtube

Aproximadamente 36,800 resultados (1.35 segundos)

Videos Educativos – CETA
ceta.zaragoza.unam.mx > edu-vids
Videos con enfoque educativo relevante a las carreras de la FES Z y temas complementarios de ... Efectiva herramienta p...

Recursos | Colegio de Ciencias y Humanidades
cch.unam.mx > recursos
SSAA VIDEOS. 5 subscribers. Subscribe · Ambiente Virtual de Idiomas. Info. Shopping. Tap to unmute. If playback does...

Aulas Virtuales – CETA
ceta.zaragoza.unam.mx > aulas-virtuales
... los estudiantes solicitan mayor uso de recursos tecnológicos como aulas virtuales o incorporación de recursos como vi...

Recursos Digitales. Alumno | Colegio de Ciencias y Humanidades
www.cch.unam.mx > recursos > alumno
Ambiente Virtual de IdiomasRefuerza tu aprendizaje en idiomas extranjeros con recursos que ofrece la CUAED (Coordin...

CETED FES Acatlán - YouTube
YouTube > CETEDFESacatlan > feed
Canal de videos del Centro Tecnológico para la Educación a Distancia de la Facultad de Estudios Superiores Acatlán de...

| Biné
bine.iztacala.unam.mx
Recuperado de: https://bine.iztacala.unam.mx/video/lausura. Resumen ... Educación y Cultura Libre, Segundo Encuentro En Línea CHAT. [Video].

Recursos Digitales. Profesor | Colegio de Ciencias y Humanidades

Sitios web en los que buscar

Añadir Eliminar Filtro Etiqueta ▾

11 - 20 de 43 < >

Sitio	Etiqueta	Disponible en la API Site Restricted JSON ?
<input type="checkbox"/> https://www.facebook.com/UNAMSa...		✗ ?
<input type="checkbox"/> https://www.facebook.com/pg/ENES...		✗ ?
<input type="checkbox"/> https://www.facebook.com/pg/PulsoC...		✗ ?
<input type="checkbox"/> https://idie.enes.unam.mx/		✗ ?
<input type="checkbox"/> https://enes.unam.mx/capacitaciondo...		✗ ?
<input type="checkbox"/> https://sites.google.com/enp.unam.m...		✗ ?
<input type="checkbox"/> https://ceta.zaragoza.unam.mx/		✗ ?
<input type="checkbox"/> https://www.youtube.com/c/EventosE...		✗ ?
<input type="checkbox"/> https://www.youtube.com/channel/U...		✗ ?
<input type="checkbox"/> https://www.youtube.com/channel/U...		✗ ?

</GRACIAS>

Mario Alberto Arredondo
malag@unam.mx

Subdirección de Visibilidad Web
DGTIC - UNAM