



Búsqueda Personalizada de Google Google Custom Search[®]

Ing. Alfredo Alonso Peña

¿PARA QUE ME SIRVE UN BUSCADOR?

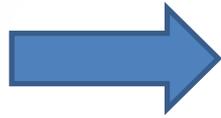
- El objetivo de un sitio Web es publicar información útil al usuario y facilitar que este usuario pueda llegar a esta información.
- Podemos utilizar algunos elementos que nos ayuden a garantizar ese acceso.
 - Tener bien organizado y estructurado nuestro sitio, ayudándonos de Mapas de Sitio, acceso al menú en todas las páginas, línea de Ariadna entre otros.
 - Proporcionar una herramienta de búsqueda de los contenidos del sitio.

¿QUÉ OPCIONES TENGO PARA IMPLEMENTAR UN BUSCADOR?

- Podemos desarrollar uno a la medida de nuestro sitio.
- En caso de usar un CMS, se puede utilizar el buscador que cada administrador proporciona.
- Usar el buscador personalizado de Google.

¿Cómo funciona?

Antes de la
búsqueda.



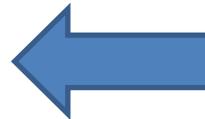
Rastrea sitios web



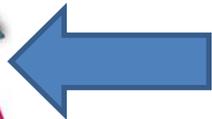
Calcula PageRank



Durante la
búsqueda.



Envía consulta



Explora índice



Selecciona documentos



Clasifica resultados



Presenta resultados

¿QUÉ PUEDO HACER?

- Crear motores de búsqueda (*colecciones de sitios o páginas*).
- Habilitar la búsqueda de imágenes en tu sitio.
- Personalizar el diseño de los resultados de búsqueda.
- Agregar resultados de búsqueda promocionados.
- Permite clasificar y restringir resultados.
- Asignar prioridades en los resultados de búsqueda.
- Uso de sinónimos.
- Priorizar el indexamiento de las páginas más importantes de tu sitio.
- Asociar un motor de búsqueda a una cuenta *AdSense* o *AdWords*.

- CUSTOM SEARCH ENGINE (*gratuito*)
- GOOGLE SITE SEARCH (*desde \$100 USD anuales*)
 - No incluir ADS
 - Cambiar el logotipo de Google
 - Recibir resultados en XML

RECURSOS PARA PROGRAMADORES

- **XML API o Custom Search API** (Permite utilizar cualquier aplicación cliente para crear, consultar y eliminar motores de búsqueda personalizados).
- **Custom Search Control** (Permite embeber un elemento de búsqueda personalizada en nuestra página web utilizando javascript).
- **JSON/ATOM API** (Permite recibir y mostrar resultados de búsqueda desde Google por medio de un Webservice REST).
100 consultas al día - \$5 USD por 1000 consultas

COLABORADORES Y REGISTRO DE AUDITORÍA

- Mediante las cuentas de administración te permite agregar colaboradores que te ayuden en la configuración de tu buscador
 - Hasta 200 colaboradores
 - Podrán ver y editar todos los datos de tu motor de búsqueda excepto: Eliminar el motor y añadir cuentas de administración.
- El registro de auditoría te mostrará las acciones que hace cada colaborador



HOLA MUNDO... DE GOOGLE

1. Logearse en Google Custom Search www.google.com/cse/
2. Llenar la información básica del motor de búsqueda
 - Nombre
 - Lenguaje
3. Indicar los sitios en los que quiero buscar
4. Aceptar los términos de uso.
5. Probar el motor de búsqueda



Search in CSE home



Custom Search

New search engine

› Edit search engine

▼ Help

- Help Center
- Help forum
- Support
- Blog
- Documentation
- Terms of Service

Send Feedback

Enter the site name and click "Create" to create a search engine for your site. [Learn more](#)

Sites to search

You can add any of the following:

Individual pages: `www.example.com/page.html`

Entire site: `www.mysite.com/*`

Parts of site: `www.example.com/docs/*` or `www.example.com/docs/`

Entire domain: `*.example.com`

Language

Name

By clicking 'Create', you agree with the [Terms of Service](#).

- En la pestaña de diseño definir la forma en que se quieren obtener los resultados.
 - Ancho Total
 - En dos páginas
 - Compacto
 - Dos columnas
 - En dos páginas
 - Solo Resultados
 - Página alojada en Google
- Elegir el los estilos para el diseño del buscador.

```
<!-- Put the following javascript before the closing </head> tag. -->
<script>
  (function() {
    var cx = '000862918480106004188:yekw5mwruco';
    var gcse = document.createElement('script'); gcse.type = 'text/javascript'; gcse.async = true;
    gcse.src = (document.location.protocol == 'https:' ? 'https:' : 'http:') +
      '://www.google.com/cse/cse.js?cx=' + cx;
    var s = document.getElementsByTagName('script')[0]; s.parentNode.insertBefore(gcse, s);
  })();
</script>

<!-- Place this tag where you want the search box to render -->
<gcse:searchbox-only></gcse:searchbox-only>
```

Pega este código en <http://>, donde quieras que aparezcan los resultados de búsqueda.

```
<!-- Put the following javascript before the closing </head> tag. -->
<script>
  (function() {
    var cx = '000862918480106004188:yekw5mwruco';
    var gcse = document.createElement('script'); gcse.type = 'text/javascript'; gcse.async = true;
    gcse.src = (document.location.protocol == 'https:' ? 'https:' : 'http:') +
      '://www.google.com/cse/cse.js?cx=' + cx;
    var s = document.getElementsByTagName('script')[0]; s.parentNode.insertBefore(gcse, s);
  })();
</script>

<!-- Place this tag where you want the search results to render -->
<gcse:searchresults-only></gcse:searchresults-only>
```

Agregar sitios Web

Puedes incluir sitios y páginas en tu motor de búsqueda de forma individual o masiva.

Páginas en tu dominio, directorios, subdominios, dominios externos.

Patrón	Descripción	Coincide
www.ejemplo.com/	Coincide con una sola página.	www.ejemplo.com/ ejemplo.com/
www.ejemplo.com/*	Coincide con todas las URL que comienzan con www.ejemplo.com o ejemplo.com.	www.ejemplo.com www.ejemplo.com/sellos ejemplo.com/sellos
www.ejemplo.com/*cometas	Coincide con todas las URL que comienzan con www.ejemplo.com/ o ejemplo.com/ y que contienen la palabra "cometas".	www.ejemplo.com/cometas.html www.ejemplo.com/cometas/pagina2.html www.ejemplo.com/diversionconcometas.html
www.ejemplo.com/producto.asp*cat=Elec	Coincide con todas las URL que comienzan con "www.ejemplo.com/producto.asp*cat=Elec".	www.ejemplo.com/producto.asp? sku=20283&cat=Elec
www.ejemplo.com/*cometas*volar	Coincide con todas las URL que comienzan con www.ejemplo.com/*cometas*volar y que contienen las palabras "cometas" y "volar".	www.ejemplo.com/cometas/como/volar.html www.ejemplo.com/volar/como/cometas.html
.ejemplo.com/	Coincide con todos los subdominios de ejemplo.com.	www.ejemplo.com/sellos host.principal.ejemplo.com/cometas ejemplo.com/cometas/volar.html

- Una promoción es una asociación entre un conjunto predefinido de términos de consulta y un enlace a una página Web.
 - Configurar consultas
 - Configurar Título y Url del resultado
 - Vigencia de la promoción
- Se salta la regla de que un documento debe estar en el índice de Google para aparecer en los resultados de búsqueda.

USO DE SINÓNIMOS

Te permite
ampliar el
espectro de
búsqueda

Panel de control - Sinónimos: UNAM 09

Puedes ampliar las consultas de búsqueda de los usuarios utilizando sinónimos, que son variantes de un término de búsqueda. Por ejemplo, puedes especificar que las búsquedas de **coches** se devolverán resultados para ambos **coches** y **automóviles**. [Más información](#)

<input type="button" value="Añadir..."/> <input type="button" value="Eliminar"/> <input type="button" value="Subir..."/> <input type="button" value="Descargar..."/> Resultados 1 - 12 de 12	
<input type="checkbox"/> Término de búsqueda	Sinónimos
<input type="checkbox"/> apoyos	becas
<input type="checkbox"/> beca	pronabes,idioma,becas unam,ecoec,dgire,Programa de Becas Chevening,conacyt,Fundación Carolina
<input type="checkbox"/> biblioteca	Biblioteca Central de la UNAM,libros,sistema bibliotecario,nacional,acervos,bibliojuridica,revista
<input type="checkbox"/> bolsa de trabajo	empleo,ofertas,curriculum,fca,fes egresados
<input type="checkbox"/> calendario	calendario anual,calendario semestral,fin de semestre,inicio de clases,vacaciones administrativa
<input type="checkbox"/> calificaciones	historial academico
<input type="checkbox"/> gaceta	gaceta en línea,E-Journal,Revista Digital Universitaria,Publica tu obra,Latindex,revistas
<input type="checkbox"/> idioma	cele,fes,facultades idiomas
<input type="checkbox"/> licenciatura	carreras,estudios,diploma,maestrias,cursos y diplomados
<input type="checkbox"/> posgrado	maestria,doctorado,especializaciones,diplomados
<input type="checkbox"/> primer ingreso	bachillerato,comipems,licenciatura,convocatoria,posgrado
<input type="checkbox"/> red inalámbrica	riu,infinitem,internet,laptop,navegar

AUTOCOMPLETAR CONSULTAS

- Aparecen cuando los usuarios empiezan a escribir en el cuadro de búsqueda
- Se puede usar los resultados promocionados en la opción de autocompletar



- **Crawlers (Rastreadores)** descubren las páginas en enlaces a otras páginas, un sitemap complementa estos datos para aquellas páginas que no están enlazadas desde ninguna otra página.
- Los sitemaps son una manera de informar a los motores de búsqueda sobre páginas que no están disponibles para su rastreo.
- **Archivo XML** Que incluye y enumera una lista de páginas de mi sitio.



- Permite agregar meta-datos:
 - Última actualización
 - Frecuencia de modificación
 - Importancia
- Enviar sitemaps es útil para poder indexar:
 - Contenido dinámico
 - Páginas web que no son fácilmente visibles para **Googlebot** (Google web crawler)
 - Indexar páginas con **AJAX** y **FLASH**
 - Páginas que no están enlazadas desde ninguna otra
 - Indexar una gran cantidad de archivos.

SITEMAPS - CONSIDERACIONES

- Cada sitemap debe contener menos de 50,000 URL's y ser menor a 10MB.
- La URL del sitemap debe tener formato UTF-8 y estar codificada.
- Especificar URL's completas
- También se pueden agrupar un conjunto de sitemaps dentro de un archivo sitemap_index.xml
- Los archivos de sitemaps son independendientes del idioma del contenido.
- El archivo sitemap debe tener el espacio de nombre XML:
`xmlns="http://www.sitemaps.org/schemas/sitemap/0.9"`

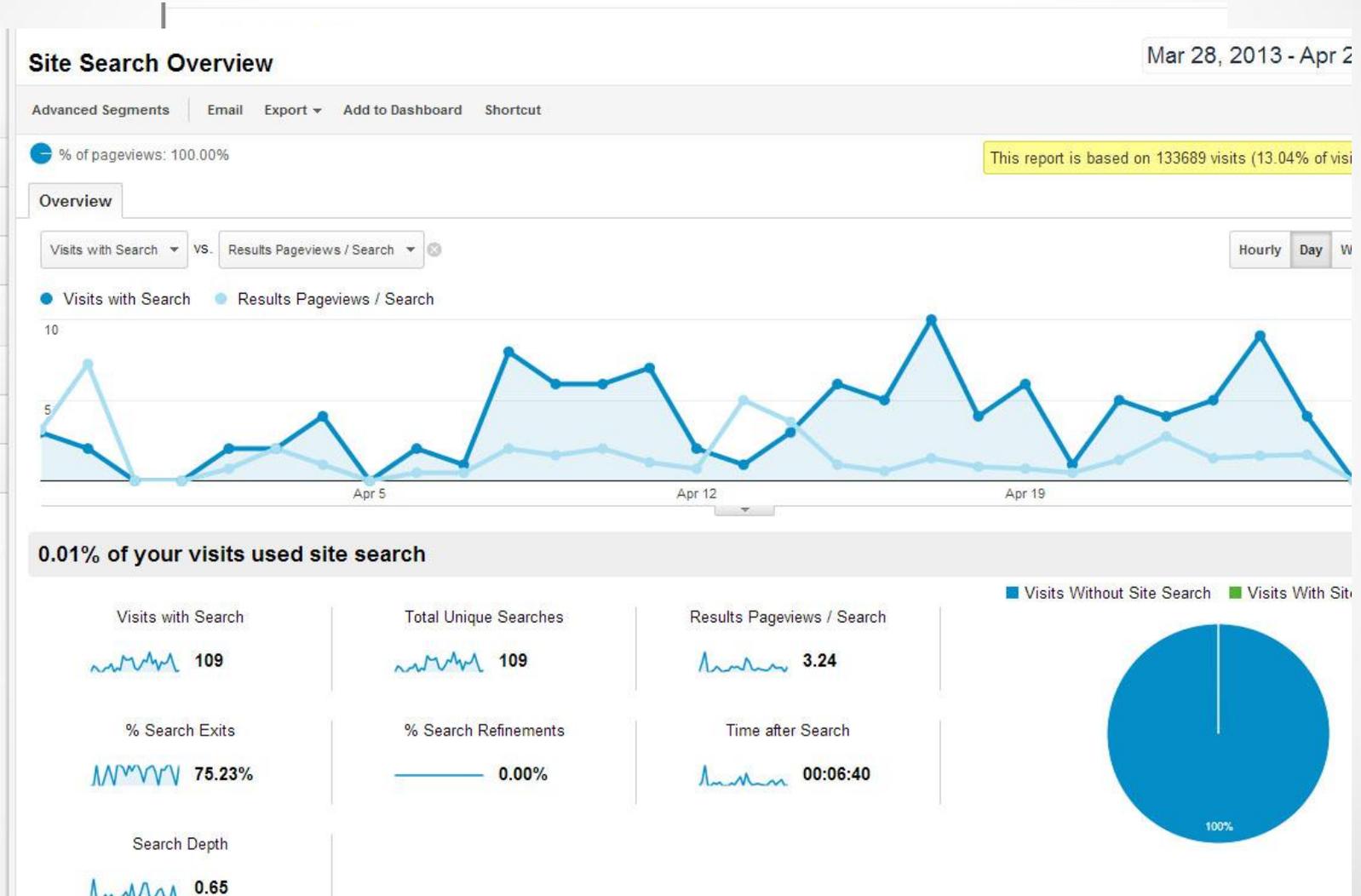
SITEMAPS - EJEMPLO

```
<?xml version="1.0" encoding="UTF-8"?>
<urlset xmlns="http://www.sitemaps.org/schemas/sitemap/0.9">
<url>
  <loc>http://www.tic.unam.mx/</loc>
  <changefreq>weekly</changefreq>
  <priority>1.00</priority>
</url>
<url>
  <loc>http://www.tic.unam.mx/directorio.html</loc>
  <changefreq>weekly</changefreq>
  <priority>0.90</priority>
</url>
<url>
  <loc>http://www.tic.unam.mx/alumnos.html</loc>
  <changefreq>weekly</changefreq>
  <priority>0.80</priority>
</url>
```

HERRAMIENTAS PARA DAR SEGUIMIENTO AL MOTOR DE BÚSQUEDA

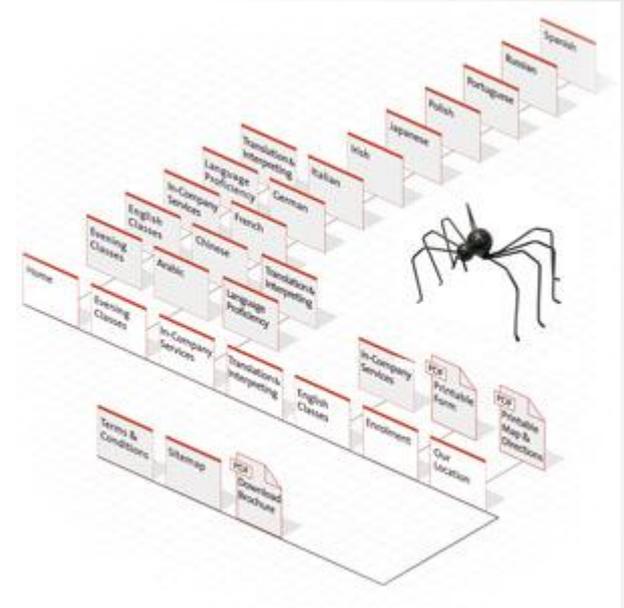
- **Estadísticas.** Permite observar el comportamiento periódico del motor de búsqueda.
 - Consultas distintas.
 - Consultas más populares
- **Google™ analytics.** Ofrecen datos y estadísticas sobre cómo los usuarios utilizan la búsqueda en tu sitio
- **Google™ Webmaster tools.** Te permite como administrador de sitios Web, saber si está definiendo de manera correcta el contenido que muestra en su sitio Web.
 - Consultas principales.
 - Páginas de destino.
 - Te permite descargar estadísticas de búsqueda

HERRAMIENTAS PARA DAR SEGUIMIENTO AL MOTOR DE BÚSQUEDA



INDEXACIÓN BAJO DEMANDA

- Hacer nuestras páginas web más visibles.
- Las páginas indexadas por Google son las que podemos visualizar en el motor de búsqueda personalizado.
- Páginas que no han pasado por el crawler de Google no estarán disponibles.
- Enviar un sitemap a Google para incluir aquellas páginas que aún no indexa.
- Google empezará a indexar registros del sitemap después de 24 hrs.



INDEXACIÓN BAJO DEMANDA

- Se puede solicitar indexación inmediata de:
 - URL's individuales
 - URL's enlazadas desde una página específica o sitemap.
 - URL's contenidas en un sitemap
- Google comenzará a indexar las páginas marcadas con mayor prioridad.
- Google solo indexa páginas a las que tiene acceso.
- Asegurarse que las páginas que se desean indexar no se encuentren en el archivo robots.txt
- Puede pedir que se indexen hasta 200 url's y eliminar hasta 20 documentos del índice de Google.

CONFIGURACIONES AVANZADAS

- **Archivo de contexto.** Describe las especificaciones del motor de búsqueda y define su comportamiento.
- **Archivo de anotaciones.** Especifican los sitios que se incluyen en el motor de búsqueda, permite agregar grandes lotes de colecciones de sitios o páginas

ARCHIVO DE CONTEXTO

```
<CustomSearchEngine
  volunteers="false"
  keywords="clima &quot;calentamiento global&quot;
  language="es-mx"
  visible="false"
  encoding="UTF-8">
<Title>Calentamiento global</Title>
<Description>Investigaciones sobre el calentamiento global.</Description>
<Context>
  <BackgroundLabels>
    <Label name="_cse_hwbuiarvsbo" mode="FILTER"/>
    <Label name="_cse_exclude_hwbuiarvsbo" mode="ELIMINATE"/>
  </BackgroundLabels>
</Context>
<LookAndFeel nonprofit="true"/>
<ImageSearchSettings enable="true" layout="1" />
</CustomSearchEngine>
```

Este documento nos permite agregar grandes lotes de colecciones de sitios o páginas para incluir en el motor de búsqueda.

Elementos:

- Site (El sitio que se desea asociar)
- Label (Define como tratar al sitio descrito)
 - Incluida
 - Excluida
 - Promocionada
 - Filtrada

ANNOTATIONS | FORMATOS SOPORTADOS | XML

- Annotations
 - Annotation
 - Label
 - Comment

```
<Annotations>  
  <Annotation about="www.unam.mx/*">  
    <Label name="_cse_Ansi-stoubiq"/>  
    <Comment>Portal UNAM</Comment>  
  </Annotation>  
  <Annotation about="www.fca.unam.mx/*">  
    <Label name="_cse_exclude_Ansi-stoubiq"/>  
  </Annotation>  
</Annotations>
```

ETIQUETAS DE MEJORA

- Son la mejor opción para clasificar y promocionar los resultados de una búsqueda.
- Permiten al usuario ejecutar una búsqueda relacionada con ciertos términos que pudieran ser de mayor utilidad o relevancia para el.
- A través del archivo de anotaciones se asocia una etiqueta a distintos sitios y a su vez se asigna una puntuación a las etiquetas

```
<Annotations>
```

```
<Annotation about="preuniversitarios.universia.net.mx/*" score="0.0001">
```

```
<Label name="planteles"/>
```

```
</Annotation>
```

```
<Annotation about="*.dgae.unam.mx/comipems/*" score="0.7">
```

```
<Label name="ingreso"/>
```

```
</Annotation>
```

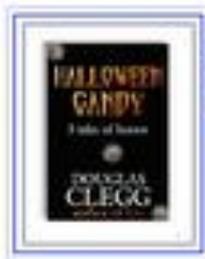
```
<Annotation about="dgenp.unam.mx/*" score="1">
```

```
<Label name="bachillerato"/>
```

```
</Annotation>
```

```
</Annotations>
```

- Formato de datos estructurados de Google.
- Nos permite agregar datos y notas en páginas web.
- Integrar acciones en el buscador personalizado según los datos.



Halloween Candy: 3 Tales of Horror by Douglas Clegg

Sep 26, 2009 ... Trick or treat! Three tales of dread and horror from award-winning novelist Douglas Clegg. Try a little **Halloween** candy this October.

www.scribd.com/doc/20258723/Halloween-Candy-3-Tales-of-Horror-by-Douglas-Clegg

[↓ Download](#) [↻ Fullscreen View](#)

Agregar PageMap en una página HTML

```
<PageMap>  
<DataObject type="action">  
<Attribute name="label" value="Descargar"/>  
<Attribute name="url"  
value="http://www.scribd.com/document_downloads/20258723?exte  
nsion=pdf"/>  
<Attribute name="class" value="download"/>  
</DataObject>  
</PageMap>
```



Gracias !!!!

Alfredo Alonso Peña

alfredo.alonso@comunidad.unam.mx