Google Custom Search Engine as a simple tool for multiple site searching

Irakli Garibashvili
Director, National Scientific Library in Georgia
Plan of a webinar

- Google Custom Search Engine (CSE) – technology insight
- How to create simple engine in 2 steps
- Multiple site search
- Searching through digital libraries (OPACs, Repositories, eJournals, etc.)
- OpenDOAR CSE
Goal

- To learn how to construct a simple search engine which will help to find information through
  - A specific site
  - Multiple sites
- And how to use the created CSE on your library website
Starting a CSE

- Go directly to a [https://cse.google.com/cse/](https://cse.google.com/cse/)
- OR search for “Google Custom Search”
- Login freely using your Google account
Click on Add button to create a first sample engine
Searching for an open course at Open University UK

Let our courses inspire you!

A new interest, a new job, a new perspective. Change your life with an Open University qualification.
Specify a URL of a site to search
Think about a proper pattern of URL – http://www.open.ac.uk/courses
Indicate a Language and Title of a CSE and just click “Create”
Result

Congratulations!
You've successfully created your Custom search engine.

Add it to your site

View it on the web

Modify your search engine
Congratulations! You have done it.

Three options here:
- Public URL – link to a search box (go there and start searching)
- Code – simple code which can be embedded to your web page and produce search from your site
- Control panel – lets you to modify or refine your engine
Control Panel
Control panel

- A control panel includes
- Tools for modification of your engine – on the left
- Demo search box, where you can make searches and test various modification options – on the right
Demo – just enter a search term and try
A result page

Bachelor of Engineering (Honours)
Modification of results layout – the left menu – “Look and feel”
Look and feel

- Gives you several options how the results will be shown:
  - On top of a search box
  - In a new browser Tab
    - Opens a new Tab and keeps your site open
    - Opens a result screen in the same Tab
  - In another page of a site
  - Etc.
- You can choose any option and test results on the right side of the Control Panel page
- When you are finished (found an option which best suits to your goals) put it you’re your website to let users make searches.
How to put a resulting CSE on your website

- Public URL – just put a link with a proper place and text
- Embed a code to your website to let users make searches directly from your site
Get code – click on a button to find the code

Get code for search box

Copy the following code, and paste it into a `<div>` element in your site’s `<body>` section, where you want the search box to render.

Note: For the most cross-browser compatibility, it is recommended that your HTML pages use a supported doctype such as `<!DOCTYPE html>`. CSS hover effects require a supported doctype.

```html
(function() {
    var cx = '0079289853178983702613Ajg8mydou_8';
    var gscse = document.createElement('script');
    gscse.type = 'text/javascript';
    gscse.async = true;
    '//cse.google.com/cse.js?cx=' + cx;
    var s = document.getElementsByTagName('script')[0];
    s.parentNode.insertBefore(gscse, s);
})();
</script>

<gcse:searchbox-only></gcse:searchbox-only>

You can customize the Search UI even more, or add per page customization by following the full documentation on CSE element.

If you want to use advanced features, Get the V1 code instead.
Embedding a code

- Just copy the code
- Note two versions of a code V1 & V2 – you have to choose a one which will work properly on your site
- The version depends on your CMS and you can just test both of them.
Sample – embedding a code to my site – to a blog entry
Embedding...
Code V2

Get V1 code for search box

Copy the following code, and paste it into a <div> element in your site's <body> section, where you want the search box to render.

Note: For the most cross-browser compatibility, it's recommended that your HTML pages use a supported doctype such as <DOCTYPE html>. CSS hover effects require a supported doctype.

```html
<div id='cse-search-form' style='width: 100%;'>Loading</div>
<script src='https://www.google.com/cse.js' type='text/javascript'></script>
<script type='text/javascript'>
google.load('search', '1', {language: 'en', style: google.loader.themes.v2_DEFAULT, google.setOnLoadCallback(function() {
  var customSearchOptions = {};
  var orderByOptions = {};
  orderByOptions['keys'] = [{label: 'Relevance', key: ''}, {label: 'Date', key: ''}];
  customSearchOptions['enableOrderBy'] = true;
  customSearchOptions['orderByOptions'] = orderByOptions;
  var customSearchControl = new google.search.CustomSearchControl('007928985317898370261_customSearchControl', [0.07928985317898370261],
  customSearchControl.setResultSize(google.search.Search.FILTERED_CSE_RESULTSET);
  var options = new google.search.DrawOptions();
  options.enableSearchboxOnly('https://www.google.com/cse?q=007928985317898370261_customSearchControl');
});
</script>
```

You can customize the look and feel even more or extend the functionality of the Custom Search.
Resulting page with a CSE search box

Test post 1

Your Blog entry has been created.

Submitted by IraKli on Mon, 01/18/2016 - 05:27.

You are watching this post, click to stop watching
Results of a search – “nano technology”

Google custom search
nano technology

Nanotechnology Engineering
www.asme.org/nanotechnology
Submit papers on application of nano-devices, nano-scale structures.

Electron Detectors
www.surface-concept.com/
for diffraction and spectroscopy nanotechnology
+99.99% MWGNT Producer
www.ntherma.com/
+99.99% Purity MWGNT Loading Hi Purity MWGNT Provider

Electrospin Nanospinner
www.inovenso.com/
High throughput electospinning machines for nanofiber production
References/Projects - Contacts

Showing results for nanotechnology
Search instead for nano technology

T356 - Engineering small worlds: micro and nano technologies ...
www.open.ac.uk/courses/modules/t356
The printed study materials are divided into three main themes together with a supporting text book as an introduction to the subject of nanotechnology.
Re-modifying your engine

[Image of a Google search results page with options for layout and search result appearance.]
Now results are in a new Tab
Adding a site for search from the CSE Control panel – “Add” button
Adding a URL

https://cse.google.com/cse/setup/basic?cx=007928985317898370261:jjgj8mydou_8

Include site individually  Include sites in bulk

http://yc.yale.edu/

- Include all pages whose address contains this URL
- Include just this specific page or URL pattern I have entered
- Dynamically extract links from this page and include them in my search engine

Advanced

Submit indexing and removal requests via Google Search Console. Learn more
Adding a URLs

- Several options:
- Including each URL separately or adding a list in bulk
- Add a page with specified address
- Add a pattern specifying a group of pages
- Extracting links from a specified page (good for online list of publications, where you can extract each of them individually)
- CSE has some Help documentation and you can just start with very basic options testing everything on your Control Panel before you put the results online
Now our CSE is searching in multiple sites indicated in the list – e.g. open courses from various Unis
Goal: CSE searching through e-journals from various publishers
Oxford journals – see a proper pattern
Adding URLs

https://cse.google.com/cse/setup/basic?cx=007928985317898370261963Axfr2xbildi

Edit included site

- Include all pages whose address contains this URL
- Include just this specific page or URL pattern I have entered
- Dynamically extract links from this page and include them in my search engine

Restrict Pages using Schema.org

Restrict pages from the above site list to only those that contain Schema.org types from the list below.

You can add up to ten (10) schema.org types to your Search Engine. Note that when you add a node, all its children automatically get included, so you do not need to add them again. For example, if you add CreativeWork, you do not need to add Book, ImageObject, VideoObject etc. separately.
A new option – you can switch off advertising from results
Adding URLs in bulk
Several tips for note

- CSE is not expanding a general Google search but limiting it to a specific site.
- CSE is producing not very accurate results – meaning that it does not eliminate entirely junk results but is minimizing results to ones which are more practical to you – eventually you have to make decision which one you choose.
- CSE cannot search in specified metadata fields (at least in the way described above), but it is searching through all metadata (unlike engines of ILSs or Digital repository software).
- CSE normally does not recognize truncations (which might be useful for specific languages in specific software systems). E.g KOHA can search for truncated terms (like phys or math), Dspace can search for phys* or math*.
Searching through Websites (meaning a combination of web pages) is more or less intuitive – you have specify a pattern carefully.

But CSE also makes possible to search through Websites produced via various databases (CMSs, ILSs, etc..)

For example you can create a search through Drupal sites or Wikipedia pages.

Or you can search through OPACs of various ILSs (not all).
You can create a search engine producing results of searches through multiple online catalogues produced via KOHA ILS (which is popular in EIFL countries and libraries)

As a result you can have on your website single search box through all KOHA OPACs in your country

All you need is just to create a standard CSE and add URLs like opac.library1.com, opac.library2.com, opac.library3.ge, etc.

We could even create a union OPAC of EIFL libraries this way.
CSE in Open repositories or Open e-journals

- CSE works well also in software systems like Dspace or Open Journal Systems, which are also popular in EIFL countries and libraries.
- In these cases the engine is searching through metadata as well as through full text files.
OpenDOAR CSE

OpenDOAR

The Directory of Open Access Repositories - OpenDOAR

OpenDOAR is an authoritative directory of academic open access repositories. Each OpenDOAR repository has been visited by project staff to check the information that is listed here. This in-depth approach does not rely on automated analysis and gives a quality-controlled list of repositories.

As well as providing a simple repository listing, OpenDOAR lets you search for repositories or search repository contents. Additionally, we provide tools and support to both repository administrators and service providers in sharing best practice and improving the quality of the repository infrastructure. Further explanation of these features is given in a project document Beyond the list.

The current directory lists repositories and allows breakdown and selection by a variety of criteria - see the Find page - which can also be viewed as statistical charts. The underlying database has been designed from the ground up to include in-depth information on each repository that can be used for search, analysis, or underpinning services like text-mining. The OpenDOAR service is being developed incrementally, developing the current service as new features are introduced. A list of Upgrades and Additions is available.

Developments will be of use to both users wishing to find original research papers and to service providers like search engines or alert services which need easy-to-use tools for developing tailored search services to suit specific user communities.

OpenDOAR is one of the SHERPA Services including RoMEO and JULIET, run by the Centre for Research Communications (CRC). Current development work is currently funded by JISC, with contributions from the CRC host organisation, the University of Nottingham.

OpenDOAR has also been identified as a key resource for the Open Access community (K B Oliver & R Swain, 2006 - PDF) and identified as the leader in repository directories in a study by Johns Hopkins University. OpenDOAR was one of the services which contributed to SHERPA being awarded the 2007 SPARC Europe Award for Outstanding Achievements in Scholarly Communications.
OpenDOAR is a Directory of Open Repositories including thousands of OA repositories of Universities, scholarly societies, etc. worldwide.

- More than hundred countries
- Several languages (English, French, Russian,..)
- It includes searching for repositories (where you can find OA repositories worldwide by your interest)
- As well as CSE for searching through all these databases
Direct Link to OpenDOAR CSE
http://www.opendoar.org/search.php
Or you can embed a code to your library website
The code can be produced from OpenDOAR CSE page with just a basic knowledge of software design

Or I can share it with you
Thank you!

Questions?

Irakli Garibashvili iger@hotmail.com