

EXT: Google Query (Data Provider)

Extension Key: googlequery

Language: en

Keywords: forAdmins, forintermediates

Copyright 2000-2010, Roberto Presedo, <typo3@cobweb.ch>

This document is published under the Open Content License
available from <http://www.opencontent.org/opl.shtml>

The content of this document is related to TYPO3
- a GNU/GPL CMS/Framework available from www.typo3.org

Table of Contents

EXT: Google Query (Data Provider).....	1	Data to retrieve (second tab).....	5
Introduction.....	3	Cache configuration (third tab).....	6
What does it do?.....	3	Mapping googleQuery results in TemplateDisplay.....	6
Questions ?.....	4	Known problems.....	7
Keeping the developers happy.....	4	To-Do list.....	8
Users manual.....	5	ChangeLog.....	9
Server Configuration (first tab).....	5		

Introduction

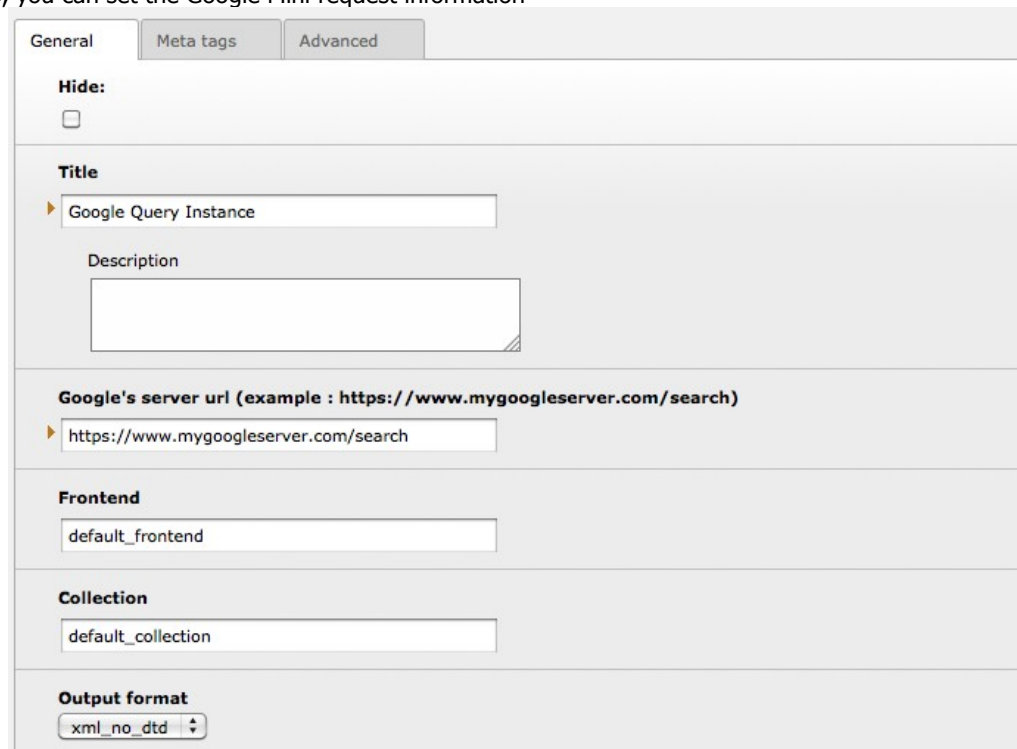
What does it do?

This extension is a Data Provider for the Tesseract Project. It sends queries to a Google server (Google Mini or Google Search Appliance) and returns the resulting data as a "recordset"-type Standardized Data Structure (see extension "tesseract" for a general explanation).

Tips for manual-editors

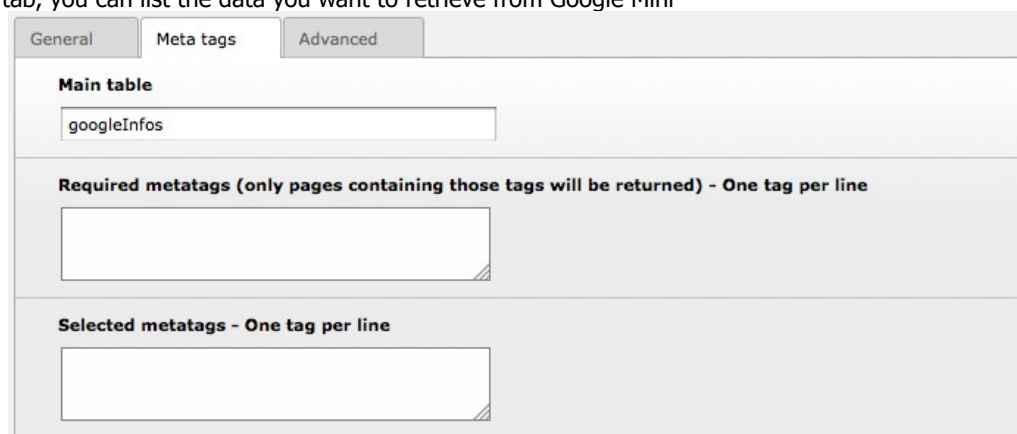
This show the typical input screen of a Google Query:

In the first tab, you can set the Google Mini request information



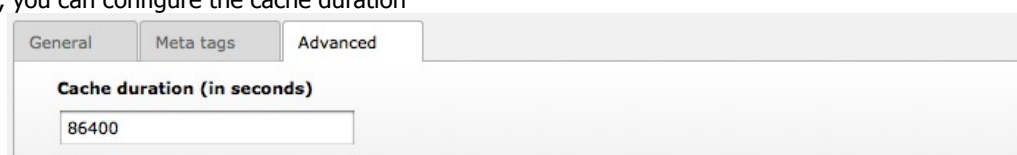
The screenshot shows the 'General' tab of the Google Query configuration interface. It features three tabs: 'General', 'Meta tags', and 'Advanced'. The 'General' tab is active. The interface includes a 'Hide:' section with a checkbox, a 'Title' section with a 'Google Query Instance' field and a 'Description' field, a 'Google's server url (example : https://www.mygoogleserver.com/search)' section with a text field containing 'https://www.mygoogleserver.com/search', a 'Frontend' section with a 'default_frontend' field, a 'Collection' section with a 'default_collection' field, and an 'Output format' section with a dropdown menu set to 'xml_no_dtd'.

In the second tab, you can list the data you want to retrieve from Google Mini



The screenshot shows the 'Meta tags' tab of the Google Query configuration interface. It features three tabs: 'General', 'Meta tags', and 'Advanced'. The 'Meta tags' tab is active. The interface includes a 'Main table' section with a 'googleInfos' field, a 'Required metatags (only pages containing those tags will be returned) - One tag per line' section with a text area, and a 'Selected metatags - One tag per line' section with a text area.

In the last tab, you can configure the cache duration



The screenshot shows the 'Advanced' tab of the Google Query configuration interface. It features three tabs: 'General', 'Meta tags', and 'Advanced'. The 'Advanced' tab is active. The interface includes a 'Cache duration (in seconds)' section with a text field containing '86400'.

Questions ?

If you have any questions about this extension, you may want to refer to the Tesseract Project web site (<http://www.typo3-tesseract.com/>) for support and tutorials. You may also ask questions in the TYPO3 English mailing list (typo3.english).

Keeping the developers happy

If you like this extension, do not hesitate to rate it. Go the Extension Repository, search for this extension, click on its title to go to the details view, then click on the "Ratings" tab and vote (you need to be logged in). Every new vote keeps the developers ticking. So just do it!

You may also take a step back and reflect about the beauty of sharing. Think about how much you are benefiting and how much yourself is giving back to the community.

Users manual

To configure a googlequery instance, you must set three types of information. First, information about the google server. Then, define the data that Google Mini's will return. And finally, the optional configuration of the cache.

Server Configuration (first tab)

– Google's server url - for example : `http://gmini.monsite.com`

As the name implies, this is the url of the Google Mini server, with `http://` or `https://` and no slash at the end.

– Frontend - by default : `default_frontend`

The name of the frontend that Google Mini must use.

In the Google Mini params, you can change the look and feel of the search and search result pages your users access. These two search pages that your users see are called a Front End. As the search and the search result pages are not going to be generated by Google Mini, but by Typo3, the setting of the output format in Google Mini is useless. However, other options can be set in Front Ends, such as KeyMatches and Related Queries settings, or remove special URLs.

By default, the Google Mini configuration comes with a Front End called `default_frontend`.

For more information about Front Ends in Google Mini :

http://code.google.com/apis/searchappliance/documentation/46/help_mini/serve_frontends.html

– Collection – by default : `default_collection`

The name of the collection that will be used by Google Mini for the request.

In Google Mini, you can create collections of documents that are subsets of the complete index. Each collection is defined by a group of URL patterns that encompasses the URLs of the documents in the collection.

– Output format – by default : `xml_no_dtd`

Here you have two options : `xml_no_dtd` (XML results or custom HTML) or `xml` (XML results with Google DTD reference). Except in special occasion, always use `xml_no_dtd`

Data to retrieve (second tab)

– Main table – by default : `googleInfos`

This is the name of the main table. This information is required, because it will be used in the Template Display extension to loop informations. By default, this value is set to `googleInfos`, as every googleQuery will at least return information about the search results (see table below).

If this was a SQL query, this information would be the FROM table.

– Required metatags

only items containing those tags will be returned. You can set as many required meta tags as you want, but enter one tag per line.

If this was a SQL query, this information would be the WHERE fields.

– Selected metatags

This is the list of every tags that could be mapped in the Template Display extension.

In addition to the list, googlequery adds a list of information about the search results.

Field name	Description
<code>googleInfos\$url</code>	The URL of the search result.
<code>googleInfos\$title</code>	The title of the search result.
<code>googleInfos\$snippet</code>	The snippet for the search result. Note: Query terms appear in bold in the results. Line breaks are included for proper text wrapping.
<code>googleInfos\$pagelang</code>	Indicates the language of the search result. The LANG element contains a two-letter language code.
<code>googleInfos\$crawldate</code>	Shows the date when the page was crawled. It is shown only for pages that have been crawled within the past two days.
<code>googleInfos\$rankpage</code>	Provides a general rating of the relevance of the search result.
<code>googleInfos\$resultnumber</code>	The index number of this search result

Field name	Description
googleInfos\$total	The estimated total number of results for the search. The estimate of the total number of results for a search can be too high or too low.

If available in the search results, informations about synonyms or keymatchs are also automatically added to the selected metatags

Field name	Description
googleSynonymes\$label	The URL of the search result.
googleSynonymes\$link	The title of the search result.
googleKeymatches\$label	The snippet for the search result. Note: Query terms appear in bold in the results. Line breaks are included for proper text wrapping.
googleKeymatches\$link	Indicates the language of the search result. The LANG element contains a two-letter language code.

If this was a SQL query, this information would be the SELECT fields.

Cache configuration (third tab)

– Cache duration

If you want to use the Tesseract cache option, just set the cache duration (number of seconds). If this is set to 0, no cache will be used.

Mapping googleQuery results in TemplateDisplay

The screenshot below shows how information provided by a googleQuery request is available in TemplateDisplay.

The screenshot shows the 'Mappings' tab in the TemplateDisplay interface. It features a 'Mapping' button and an 'HTML' button. The main area displays a code editor with the following HTML structure:

```
<h2><Search results</h2>
<ul>
<!--LOOP(googleInfos)-->
<li>
<a href="###FIELD.url###">###FIELD.title###</a><br />
<span class="snippet">###FIELD.snippet###</span><br/>
<span class="url">###FIELD.url###</span><br/>
</li>
<!--ENDLOOP-->
</ul>
```

On the right side, there is a 'Fields' dropdown menu. The dropdown is open, showing a list of fields grouped by category:

- GOOGLEINFOS**
 - url
 - title
 - snippet
 - pagelang
 - crawldate
 - rankpage
 - resultnumber
 - total
- GOOGLESYNONYMES**
 - label
 - link
- GOOGLEKEYMATCHES**
 - label
 - link

At the bottom of the interface, there is a 'Save field configuration' button.

Known problems

- Google Mini has some limitations. One is a limit of 100 maximum results returned. This means that if you use the Pagebrowse option in Template Display, the pagination may be wrong when displaying the results around the 100th position.

If you have any other issue, please refer to the Tesseract Project web site (<http://www.typo3-tesseract.com/>). You may also post your problems to the TYPO3 English mailing list (typo3.english), so that others may benefit from the answers too. For bugs or feature requests, please open an entry in the extension's bug tracker on Forge (<http://forge.typo3.org/projects/extension-googlequery/issues>).

To-Do list

The roadmap for the evolution of this extension can be found on Forge: <http://forge.typo3.org/projects/roadmap/extension-googlequery>

For feature requests, please open a report on Forge issue tracker: <http://forge.typo3.org/projects/extension-googlequery/issues>

ChangeLog

Version	Changes:
1.0.0	Initial public release