



DRAFT

**User Guide**  
**Site Search ONE Premium WordPress Plugin**

Version

1.1.0

[support@dtsearch.co.uk](mailto:support@dtsearch.co.uk)

# Site Search ONE WordPress Plugin

## Premium Add-on

Updated 19 Oct 2022

### Contents

Introduction.....	2
User Defined Synonyms.....	3
Media Library files.....	4
WP Media Category Management (WP MCM) plugin.....	4
Media Library Assistant (MLA).....	4
PDF Support.....	5
PDF hit-highlighting.....	5
PDF 1.x and PDF 2.0.....	5
PDF Fast Web View.....	5
Encrypted PDF files.....	5
PDF Metadata.....	5
Search Results format for PDF files.....	5
Legal PDF Demo.....	6
Mimetype filter.....	6
PDF Meta data filters.....	6
Setting up PDF Attachment Custom Fields.....	7
Prerequisites.....	7
Setup.....	7
Setup Advanced Custom Fields (ACF):.....	8
Setup Media Library Assistant to extract PDF Metadata:.....	9
Verify that field data was extracted.....	10
Index the documents.....	10
Search Analytics.....	11
WP Site Insights plugin.....	11
WP Search Analytics plugin.....	11
Other add on features.....	12
Get Help.....	13
Feedback.....	13

# Introduction

Site Search ONE WordPress Plugin is designed to provide professional grade search for WordPress sites that have high volumes of Posts and Pages. The Premium add-on is required for features that extend the basic functionality.

We will continue to improve the basic version and will also offer premium addons for sites that need more features. This guide covers premium features that are available now or in beta test stage.

## IMPORTANT NOTES ON BETA BUILDS

No automatic updating is available for beta versions; when a new beta version is released, deactivate the currently installed version, then delete it from your website, before uploading and activating the latest build. We will advise you by email when updates to the plugin or documentation or service occur.

We recommend that you test it on a local server or staging site, before installing on a live site.

During beta phase the design may change, backward compatibility will be preserved where possible, but cannot be guaranteed.

## FEATURE REQUESTS & BUG REPORTING

Send all bug reports, and requests for new features or changes to [support@dtsearch.co.uk](mailto:support@dtsearch.co.uk) with as much detail as possible.

We will add all requests for features to a wish list; if your application cannot be satisfied by the current release, please let us know what features are needed urgently and we will try to prioritize your request.

**Our aim is to offer you the best 'no code' site search solution for your WordPress CMS site!**

**Thank you for your co-operation!**

# User Defined Synonyms

The basic Site Search ONE Plugin is supplied with synonyms using the English WordNet. If your site content is not in English or you need synonyms that are specific to your market or site content this add-on is the solution.

A subscription to the **User Thesaurus Plus** add-on is included in the Premium subscription. User Thesaurus Plus allows you to create multiple synonym files, these can be assigned to a particular search page or index.

The screenshot shows the configuration interface for the 'Indexes' section. It includes several sections: 'Indexes' with a 'Search all' checkbox and a list of search categories (Book Search, News Search, Legal, Products & News, Products and News); 'Index name' with a text input field containing 'Car Search'; 'Search bar' with checkboxes for 'Hide search bar' and 'Hide Stemming and Synonyms Checkboxes'; 'Synonyms' with a checked 'User Thesaurus Plus' checkbox, a file selection area showing 'UT\_sample\_colours.xml', and a checked 'WordNet' checkbox; and 'Search type' with a table of search options.

Type	Default	User Choice
Any words	<input type="radio"/>	<input checked="" type="checkbox"/>
All words	<input type="radio"/>	<input checked="" type="checkbox"/>
Boolean (and, or, not, ...)	<input checked="" type="radio"/>	<input checked="" type="checkbox"/>

Save

## Synonyms

If you want to define your own synonyms instead of WordNet synonyms, select **User Thesaurus Plus**, then press **Choose File** to select one of the xml files generated by **User Thesaurus Plus**.

If you want to use both, check the **WordNet** checkbox also.

In the example shown the `UT_sample_colours.xml` file supplied with User Thesaurus Plus has been applied to the Car Search index, so that a search for scarlet or crimson for example will return results for 'red' cars.

Note that the recommended maximum size for a user defined synonym file is 10,000 items, the speed of searches will be slower with larger synonym sets.

## Media Library files

From version 1.9.0 of the Basic plugin, when the Premium plugin version 1.1.0+1087 or later is installed it enables Site Search ONE to search items uploaded to the WordPress Media Library. It has Admin options to search Media that are attached or unattached to Posts, options to include Media in certain Media Categories only, and option to choose selected MIME types only (e.g., PDF, Excel, Word).

When the Premium plugin is installed, it will add extra Media options to the Site Search ONE **Search Page > Edit** screens.

Note: Site Search ONE will skip indexing any documents greater than 20mB.

## WP Media Category Management (WP MCM) plugin

SiteSearchONE is compatible with this plugin, if installed you can use the MCM Category as a Facet.

## Media Library Assistant (MLA)

The MLA plugin extracts metadata from PDF and other files. We recommend this plugin in combination with the Advanced Custom Fields (ACF) plugin for creating indexable PDF fields in Site Search ONE.

# PDF Support

## PDF hit-highlighting

The PDF files should be in "image with hidden text" format to support hit highlighting. Site Search ONE uses a server-based hit-highlighter and does NOT require you to install the dtSearch Hit-highlighter on your browser.

## PDF 1.x and PDF 2.0

PDF 1.x and PDF 2.0 files are both supported.

## PDF Fast Web View

Adobe Acrobat can restructure PDF files so they can open more quickly in a web browser, this is called Web Enabled. This Adobe article explains how it works:

<https://helpx.adobe.com/acrobat/using/optimizing-pdfs-acrobat-pro.html>

Using the Action Wizard in Adobe Acrobat, you can optimize an entire folder of PDF files in one step: <https://helpx.adobe.com/acrobat/using/action-wizard-acrobat-pro.html>

Ref: <https://support.dtsearch.com/DTS0116.htm>

## Encrypted PDF files

Encrypted PDF files cannot be indexed unless the PDF file can be opened without a password and the PDF file permissions allow for text extraction.

## PDF Metadata

PDF files have several standard fields such as Title, Author, Published Date, they can also contain Custom Fields such as Case Title or Date of Judgment as well as IPTC/EXIF metadata.

Ref: <https://support.dtsearch.com/DTS0141.htm>

## Search Results format for PDF files

For PDF files, the display name in search results uses the Title field as supplied in either the post editor or WordPress Dashboard>Media>Edit window.

## Legal PDF Demo

From the search menu (top right) at [dtsearch.co.uk](http://dtsearch.co.uk) select Categorized search, select Legal PDF from the drop down on the left of the search bar.

If you click on the **filter**  button on the search bar, you will see that search results can be filtered by **Date of Judgement** or **Case Title**, as well as by **Mimetype**.

### Mimetype filter

The main MIME media type for PDF is `application/pdf`; you can limit the search results to PDF documents by choosing that Mimetype filter. The demo also shows that you can limit search results by using the facet checkbox **MCM Category** (this demo uses the Media Category Management plugin).

### PDF Metadata filters

The files in the Legal PDF demo have Custom fields of **Date of Judgement** and **Case Title**. You can use Adobe Acrobat or other PDF editing applications to create Custom fields in your PDF files before you upload them to Site Search ONE.

To enable Search Site ONE to index these fields we recommend using the **Media Library Assistant** and **Advanced Custom Fields** plugins. [See the next section.](#)

# Setting up PDF Attachment Custom Fields

How to use **Media Library Assistant** and **Advanced Custom Fields** plugins to extract fields from PDF documents for use as text or date filters. The Legal PDF demo files are used as an example.

## Prerequisites

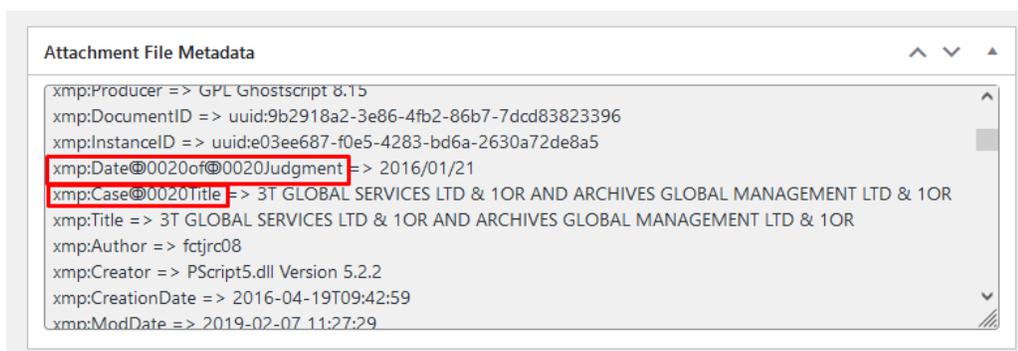
Install the following plugins:

1. **Media Library Assistant** (<https://wordpress.org/plugins/media-library-assistant/>)
2. **Advanced Custom Fields** (ACF) (<https://wordpress.org/plugins/advanced-custom-fields/>)

Before uploading PDF files, check they all have the same required fields.

## Setup

1. All PDF documents should be assigned a common Media Category. You will need a plugin to achieve this, since WordPress can only assign Categories to Posts and Pages by default. Our Legal PDF demo uses the Media Category Manager (MCM) plugin, but others such as **Media Library Organizer** can add a Media Category option to WordPress.
2. Find the names of the fields to be extracted from PDF files:
  - a. In WordPress backend, go to **Media** in the left-hand pane.
  - b. Select a PDF file that contains the fields to be extracted.
  - c. In the **Edit Media** window scroll down to the **Attachment File Metadata** block. The field names found in the PDF are listed (everything to the left of the => sign):



For the Legal PDF demo, the required fields are

xmp:**Date@0020of@0020Judgment**

xmp:**Case@0020Title**

Make a note of the field names, and the format of the date (2016/01/21 = Y/m/d)

**Note:** Use fields starting **xmp:** (Not pdf:)

## Setup Advanced Custom Fields (ACF):

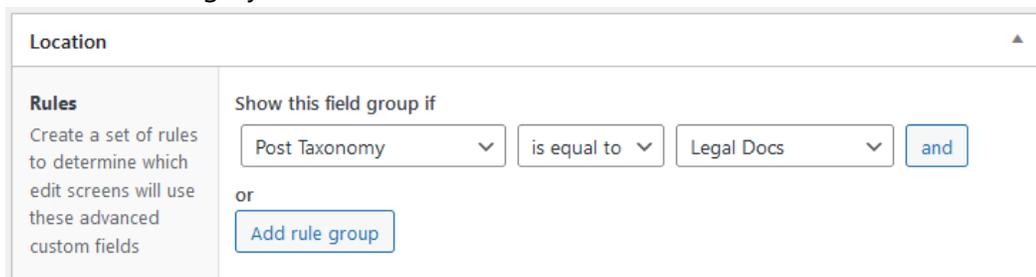
In WordPress backend, go to **Custom Fields** in the left-hand pane

1. In **Field Groups** click **Add New**



2. Supply a **Title** e.g., 'Legal Document PDF Fields'
3. To add a **text** field:
  - a. Click **Add Field**
  - b. Enter a **Field Label** e.g., 'Case Title'  
The field label is what will be displayed in Site Search ONE.  
It does not need to match the PDF field.
  - c. Set **Field Type** to **Text**
  - d. Click the **Field Name** textbox – it will be populated automatically with a value
  - e. e.g., 'case\_title' – Make a note of this name.
  - f. Leave everything else at default/blank. Click **Close Field**.

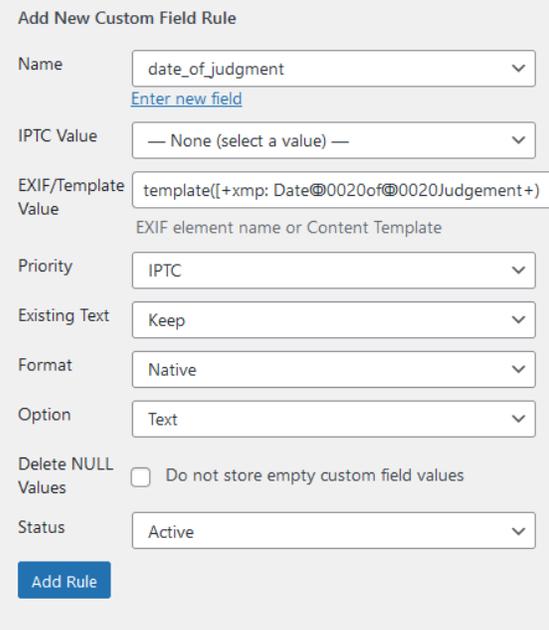
- 
4. To add a **date** field:
    - a. Click **Add Field**
    - b. Enter a **Field Label** e.g., 'Date of Judgment'
    - c. Set **Field Type** to **Date Picker**
    - d. Set **Display Format** and **Return Format** to **Custom**.
      - I. Set the formats to match the format of the date in the PDF documents using WordPress date format strings, see this page:  
<https://wordpress.org/support/article/formatting-date-and-time/>  
For example, the date '2016/05/10' should have the format string **Y/m/d**.
      - II. Make a note of the generated **Field Name** e.g.; date\_of\_judgment
      - III. Click **Close Field**
  5. In the **Rules** section, configure **Show this field group if** the **Post Taxonomy** is equal to the Media Category of the PDF documents.

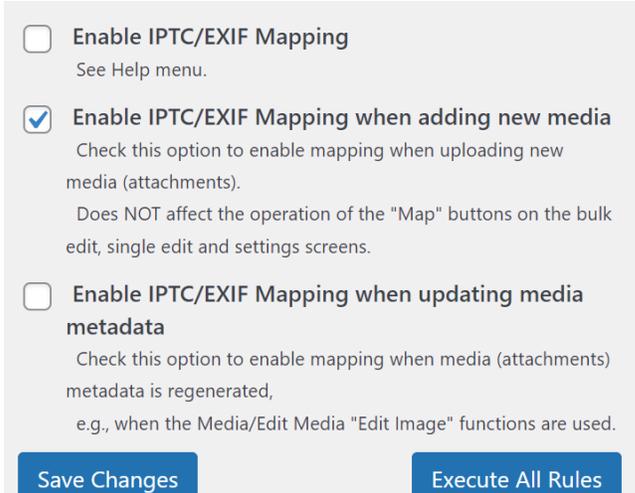
A screenshot of the WordPress Custom Fields 'Rules' section. The 'Show this field group if' section is configured with 'Post Taxonomy' selected in the first dropdown, 'is equal to' in the second, and 'Legal Docs' in the third. An 'and' button is visible between the third dropdown and the 'Add rule group' button.

6. Finally, click **Publish** to save the new field group.

## Setup Media Library Assistant to extract PDF Metadata:

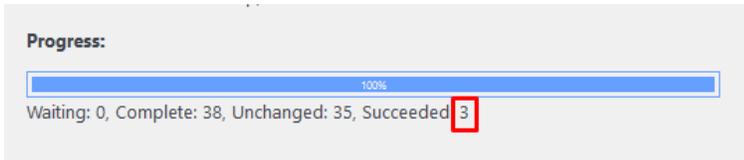
- 1) In the WordPress backend go to **Settings > Media Library Assistant**
- 2) Select the **IPTC/EXIF** tab

	<p>Under <b>Add New Custom Field Rule</b> in the left-hand pane:</p> <ol style="list-style-type: none"><li>a) <b>Name:</b> select one of the field names we created in ACF e.g.; <b><i>case_title</i></b> or <b><i>date_of_judgment</i></b></li><li>b) Leave <b>IPTC Value</b> as <i>None</i></li><li>c) <b>EXIF/Template Value:</b> enter the PDF field names we noted earlier: template:([+xmp:&lt;<b>FIELD NAME</b>&gt; +])</li></ol> <p>Replace &lt;<b>FIELD NAME</b>&gt; with the desired field, in our example: template:([+xmp:Case@0020Title+]) or template:([+xmp:Date@0020of@0020Judgement+])</p> <ol style="list-style-type: none"><li>d) Leave all other values at default.</li><li>e) Click <b>Add Rule</b>.</li><li>f) Repeat for the other field(s)</li></ol>
---	--

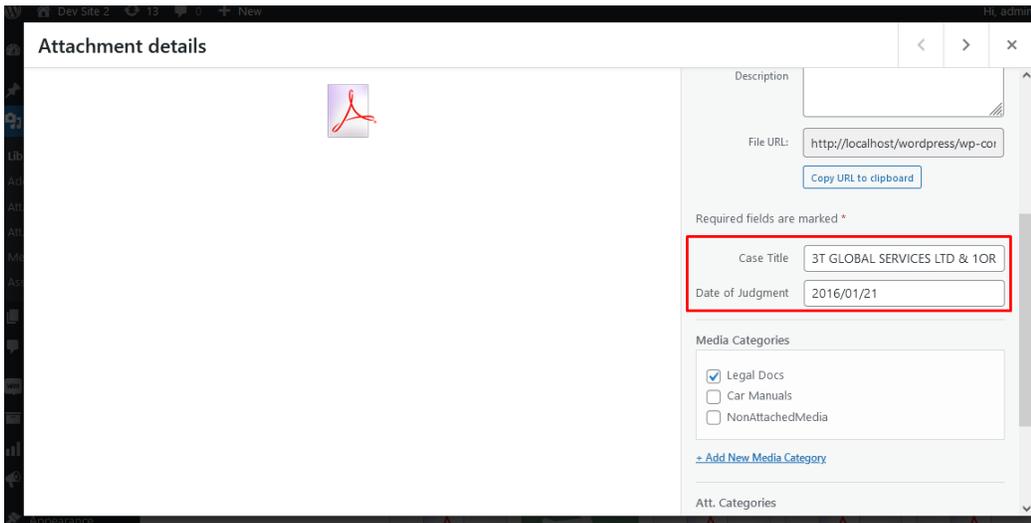
	<p>Check <b>Enable IPTC/EXIF Mapping when adding new media</b></p> <p>Click <b>Save Changes</b> This will ensure when new PDF files are uploaded that the fields are extracted.</p> <p>Click <b>Execute All Rules</b> This will extract the metadata from PDF documents already uploaded.</p>
---	---

## Verify that field data was extracted

After executing the rules, **Media Library Assistant** should report the number of documents from which it extracted the field data:



When editing a PDF file in the Media Library, verify the field data is displayed as expected:



## Index the documents

Site Search ONE will automatically detect the fields and display them as text and date filters in the filter panel after they are indexed.

New search pages will detect the fields in PDF documents.

For existing search pages, from WordPress backend **Site Search ONE > All Search Pages** click on **Rebuild** for any search page(s) that display **Index all Media belonging** to the PDF Category.

## Search Analytics

The plugins below work with Site Search ONE search bar widget or the WordPress search text box. Other plugins will not work with the Site Search ONE widget.

We recommend that you only activate one of these plugins on your site.

### WP Site Insights plugin

Details of this plugin can be found here: <https://wordpress.org/plugins/wp-search-insights/>

This plugin is free.

### WP Search Analytics plugin

Details of this plugin can be found here: <https://www.cornelraiu.com/mwt-search-analytics/> and here: <https://wordpress.org/plugins/search-analytics/>

This plugin is free.

[Contact us](#) if you prefer other analytics plugins.

## Other add on features

The dtSearch Engine has many features which may not yet be available in our REST API.

Other addon features are in various stages of development.

Our aim is to provide 'no-code' solutions so that you can get to market fast!

### **If your site needs additional features, please ask!**

We may be able to offer it in the basic version or as a premium addon.

Our REST API currently supports natural language options (All words, Any Words, Exact match) as well as Phonic and Fuzzy search options, but Phonic and Fuzzy are not currently supported in the UI backend (administrator access) or UI front end (site user access). They are available in the search query the same as in the dtSearch Desktop or dtSearch Web products.

Any search feature in dtSearch Desktop can be supplied in addition to more advanced features that the dtSearch Engine can handle such as regular expressions, email, credit-card, date matching.

Other features you might need for an eCommerce site are term or field weightings, search suggestions, search as you type, search logging, page promotion, please ask!

### **Languages**

Our basic plugin user interface is in English, it can be translated into any language supported by Unicode. In addition, if you need stemming in a language other than English, please ask, we can change your stemming language at no extra cost! Multiple stemming languages for multi-lingual sites can also be supplied at an additional charge.

### **Need more?**

AI, ML, search logging, document meta data extraction and load, back-end search (comments, enquiries, support), please ask!

## Get Help

If you need assistance with setting up Site Search ONE, please contact [support@dtsearch.co.uk](mailto:support@dtsearch.co.uk) or use the [Contact Us](#) form on the About menu or in the footer section of the dtSearch UK website.

## Feedback

Please give us feedback on this user guide so that we can provide content that is useful and helpful.  
Thank you!

**All suggestions for improvement of the product and this manual are very welcome!**