

The [smart search](#) allows users to search through the content of files uploaded as [page attachments](#). The attachment search supports both types of [file storage](#) provided by Kentico (database or file system).



The attachment search only works for files that are connected to pages through one of the following methods:

- Attachment files added to pages through fields with the **Data type** set to **File** or **Attachments** in the [page type definition](#)
- Attachments uploaded in the **Pages** application on the **Properties -> Attachments** tab of pages

## Choosing which file types are searchable

By default, the attachment search supports the following file types:

- txt
- csv
- pdf
- docx
- xlsx
- pptx
- xml
- html
- htm



**Note:** The search does NOT work for:

- Legacy MS Office formats: doc, xls, ppt
- Certain types of PDF files, including:
  - Encrypted files
  - Files using PDF version 1.5 or older

You can limit which of the file types are searchable for individual websites:

1. Open the **Settings** application.
2. Navigate to the **System -> Search** category.
3. Select file extensions in the **Allowed attachment file types** setting (if no checkboxes are selected, all supported types of attachment files are indexed).
4. Click **Save**.

If you wish to search one of the unsupported file types, you need to implement a [custom search text extractor](#).

## Enabling indexing for page attachments

The attachment search functionality is supported by indexes of the **Pages** type, for both [Azure Search indexes](#) and [local indexes](#). The attachment search is NOT available for local indexes of the **Pages crawler** type, which directly index the HTML output of pages.

To set up the attachment search for your website:

1. Open the **Smart search** application.
2. Create or edit a **Pages** type search index (either as an [Azure Search index](#) or a [local index](#)).
3. When defining the search content on the **Indexed content** tab, select the **Include attachment content** option for the index's allowed content.
4. Click **Save**.
5. Switch to the **General** tab and **Rebuild** the index.

While building the page index, the smart search processes the allowed pages, extracts the text of any attachment files and includes it in the content of the index (along with the other page data). When users perform a search using the index, the system returns results for pages whose attachments match the search expression.



#### Updating the search content of attachments (Upgrades and Hotfixes)

Kentico stores the text content extracted from page attachments in the database. When rebuilding page indexes, the search loads the "cached" attachment text from the database. The system only processes the file text directly for attachments that do not have any search content saved.

If you apply a hotfix or upgrade that changes how the search indexes attachment files, you need to clear the attachment search content:

1. Open the **System** application.
2. Select the **Files -> Attachments** tab.
3. Click **Clear attachment search cache**.

You can then **Rebuild** your page indexes, which updates the attachment content according to the new functionality.

## Configuring the attachment search

You can adjust how the system indexes attachment files by adding keys to the **appSettings** section of your application's **web.config** file.



The indexed content always includes:

- File metadata (title, tags, author name etc.)
- Comments (for example in MS Office files)

### Limiting the maximum size of indexed files

Indexing of very large files can be resource intensive and have a negative impact on your website's performance. To prevent the system from indexing files larger than a certain size, add the **CMSSearchMaxAttachmentSize** key:

```
<add key="CMSSearchMaxAttachmentSize" value="10000" />
```

The key sets the maximum allowed file size in kB. The search ignores page attachments whose size exceeds the value.

### Indexing of XML content

When indexing the content of XML files, the search does NOT include the following content by default:

- Comments
- The values of tag attributes

You can enable indexing for such content by adding the following web.config keys:

```
<add key="CMSSearchIndexXmlComments" value="true" />
<add key="CMSSearchIndexXmlAttributes" value="true" />
```

### Enabling character encoding detection for text files

By default, the search can read text files (**txt** and **csv**) that use the following character encoding:



- [UTF-8](#)
- The default Windows encoding (the operating system's current ANSI code page)

If you encounter problems when indexing text files with a different encoding type, you can enable automatic encoding detection:

```
<add key="CMSSearchDetectTextEncoding" value="true" />
```

The system then attempts to detect the encoding type for each file, and use the correct option when reading the content during the indexing process.



**Note:** Correct encoding detection is not guaranteed for all files. Automatic detection also slightly increases the time required to index text files.