



# PDF content index updates

Last Modified on 03/06/2024 11:25 am EST



Thanks to the code upgrades we made a few weeks ago, we've been able to revisit an area of [search](#) functionality that is near and dear to some of you: surfacing content within PDFs in search results.

First, some background:

When you upload a PDF to KnowledgeOwl and add it to an article (either as a [link](#) or [embedded in an iframe](#)), we run a PDF scraping tool when you save the article. That scraping tool tries to pull all the text out of the PDF so it can be included in the search index for that article.

It's not an exact science; formatting and character encoding in PDFs can prevent or garble the scraping process, but for those of you leveraging PDFs, we do it to try to make more of your content available in search.

By default, we do this for PDFs under 100 pages. If you have PDFs longer than 100 pages that you'd like indexed for search, there's an [option in Settings > Search](#) you'll need to enable.

We've rolled out one important bug fix for these long PDFs, as well as new functionality for customers using our Secure File Library:

## Bug fix for longer PDFs

If you are already using that setting, good news: we had a bug with indexing these larger PDFs, where newly-added PDFs were being scraped for search indexing but re-adding a PDF already in your File Library was not.

We've fixed that, so all PDFs regardless of age should be indexed.

If you have the index large PDFs option enabled and you have any instances of longer PDFs not showing up in search, either the PDF is formatted in a way that our scraping tool can't extract the text (sad, but it happens), OR we need to reindex your knowledge base's search results now that this bug fix is out.

If you think you fall into this category, [contact us](#) and request a search reindex at a quiet time for your knowledge base (it will knock out search results while it runs, so we don't want to do it at your busiest time!).

## Secure file library PDFs now indexed!

By default, all files stored in KnowledgeOwl can be accessed by URL, even if someone isn't logged into your knowledge base. You can override this behavior and [require authentication to view any image or file from your file library](#) (an option in **Settings > Security**). We call this our Secure File Library. If this setting is enabled, it basically means that if you send someone the URL for a file stored in KnowledgeOwl, like a PDF or an Excel sheet, they will have to login to your knowledge base before they can view it. It's an extra layer of security a few of our customers appreciate.

The one downside to using this setting is that, historically, using Secure File Library prevented our search indexing tool from scraping those PDFs.

Until now. 😞

Thanks to a release earlier this week, customers who are using Secure File Library will now start to see PDF content indexed for search! 📄

If you already have Secure File Library enabled, you'll need to trigger a reindex of the content in order to see the changes take effect. You can:

- Save a small change to an article containing a PDF. Depending on the length of the PDF, it can take a few minutes to fully scrape the contents and add it to the index, so this is a good one to do before you start making a cup of tea or as a meeting is starting. This will update the article in question but leave the rest of your knowledge base alone--it's a good test to be sure you're seeing what you expect.
- Trigger a full search reindex on your knowledge base (by adding, removing, or editing a [synonym](#), or changing other key search settings).
- Reaching out to our support owls to request that we reindex your knowledge base for search.

The full search reindex of options 2 and 3 will reindex all of your articles' content, though it can knock out your search results while it runs. If you have the setting enabled to index PDFs 100+ pages long, it can take a while to see all that PDF content show up in search. But it will ultimately ensure that all your PDF content is indexed for search.

For those of you including PDF content in your knowledge base, we hope that these changes help you get more out of our search functionality!