

# **dtSearch PDF Search Highlighter**

---



# Table of Contents

|  |   |
|--|---|
| Overview .....   | 1 |
| Compatibility and usage .....  | 1 |
| dtSearch Desktop/Network Users .....                                       | 1 |
| Web sites .....  | 1 |
| Keyboard shortcuts .....   | 1 |
| Toolbar buttons .....  | 1 |
| Options .....  | 3 |
| Which web sites should be allowed to highlight hits on PDF files?.....     | 3 |
| Verify that highlighting requests are only sent to compatible servers..... | 3 |
| Check for Updates .....  | 4 |
| Technical Information .....  | 5 |
| How the plug-in works .....  | 5 |
| Overview.....  | 5 |
| Installing the plug-in .....   | 5 |
| How hit highlighting data is included in URLs.....                         | 6 |
| Security Options.....  | 6 |
| Servers that can generate hit highlighting information .....               | 6 |
| Validation of highlighting script.....                                     | 6 |
| Information for Developers .....   | 7 |
| Internet Explorer Compatibility .....                                      | 7 |
| Changes needed to web sites that use dtSearch Web.....                     | 7 |
| Changes needed to dtSearch Publish applications.....                       | 7 |
| Changes needed to existing server-based applications.....                  | 7 |
| Changes needed to existing desktop applications.....                       | 8 |
| Distribution of the plug-in.....   | 8 |



# Overview

The dtSearch PDF Search Highlighter is a free plug-in for Adobe Reader that makes it possible for dtSearch and other search programs to highlight hits in retrieved documents inside Adobe Reader X. Earlier versions of Adobe Reader had this feature built in, so there was no need for a plug-in.

Subject to the included license agreement, you can download the plug-in at no charge from the dtSearch web site [here](#) and install the plug-on on any number of computers within your organization. The plug-in may be used with any compatible product or web site -- not just dtSearch products or products that use the dtSearch Engine.

## Compatibility and usage

### dtSearch Desktop/Network Users

To use the plug-in with dtSearch, you will need dtSearch version 7.67 or later. dtSearch will automatically detect the plug-in if it is installed.

### Web sites

The plug-in works with web sites to highlight hits in PDF files after a search. If a web site that you use to search PDF files has been updated to work with this plug-in, then you can just install the plug-in and leave the default settings in place. Web sites that use dtSearch Web 7.67 or later will automatically work with the plug-in.

If a web site has not been updated to work with this plug-in, then you will need to uncheck the box in [Options](#) to "Verify that highlighting requests are only sent to compatible servers."

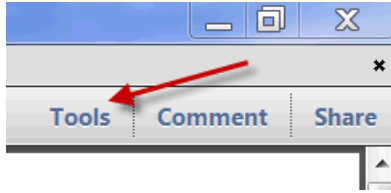
### Keyboard shortcuts

When a PDF file opens after a search, press Ctrl+Shift+Spacebar to go to the next hit and Ctrl+Shift+Backspace to go to the previous hit.

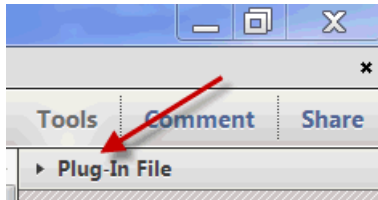
### Toolbar buttons

To make the Next Hit and Previous Hit buttons visible in Adobe Reader:

1. If the Adobe Reader toolbar is not visible, press F8 to make it visible. (To make the toolbar visible by default, open Adobe Reader and click Edit > Preferences > Internet and un-check the option to "Display in Read Mode by default".)
2. Click the "Tools" button in the Adobe Reader toolbar to open the Tools panel



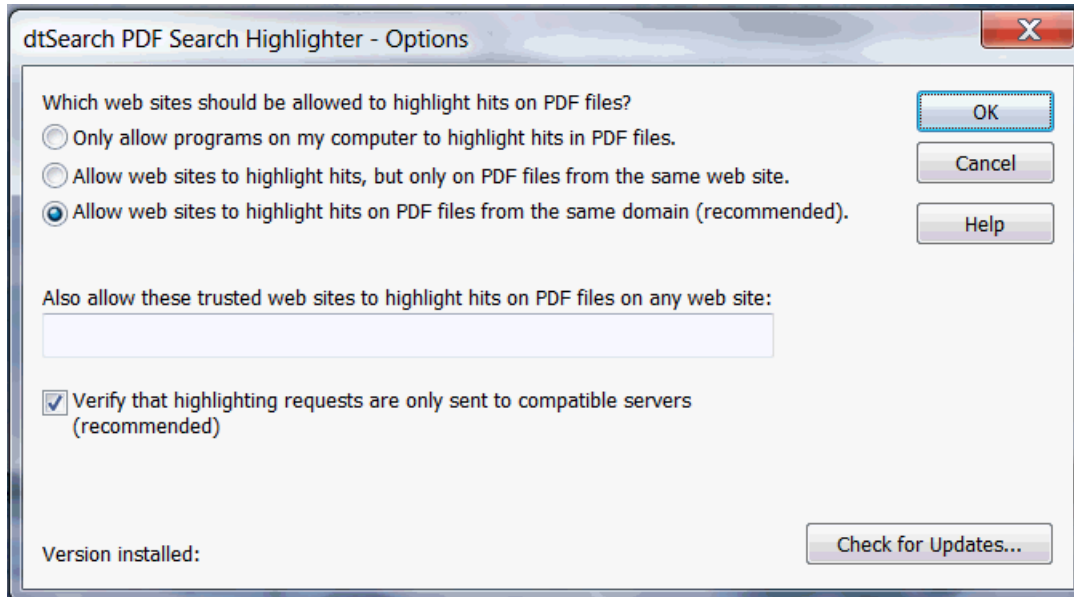
3. Click "Plug-in File" to show the Next Hit and Previous Hit buttons.



Note: Adobe Reader no longer allows plug-ins to add buttons to the main toolbar, so the plug-in can only add buttons to this Tools panel in Adobe Reader. Also, at this time, due to a bug in Adobe Reader X, only the top half of the plug-in buttons will appear if you exit and re-enter Adobe Reader without closing the "Plug-in File" section of the Tools panel. dtSearch Corp. has reported this bug to Adobe.

# Options

After installing the plug-in, click **Start > All Programs > dtSearch > dtSearch PDF Search Highlighter > dtSearch PDF Search Highlighter Options** to access security and other options for the plug-in.



These options only affect highlighting of hits in Adobe Reader and will have no effect on internet search engines such as Google and Bing that do not do this.

## Which web sites should be allowed to highlight hits on PDF files?

### **Only allow programs on my computer to highlight hits in PDF files.**

Choose this option if you only use dtSearch Desktop/Network or other desktop search products and do not want to use web sites to highlight search hits PDF files.

### **Allow web sites to highlight hits, but only on PDF files from the same web site.**

Choose this option if you use a web site to search PDF files and the PDF files are located on that same web site.

### **Allow web sites to highlight hits on PDF files from the same domain (recommended).**

Choose this option if you use a web site to search PDF files and the PDF files are located on a different web site but within the same domain. For example, you might use a search engine at <http://search.example.com> to search PDF files at <http://www.example.com>.

### **Also allow these trusted web sites to highlight hits on PDF files on any web site.**

This option lets you list web sites that you trust to provide highlight information on other sites' PDF files. For example, you might go to <http://search.example.com> to search legal opinions that are on unrelated web sites.

## Verify that highlighting requests are only sent to compatible servers

## dtSearch PDF Search Highlighter

Check this box to enable a security enhancement that verifies that highlight requests are only sent to servers that are compatible with this plug-in. For more information on the purpose of this, please see "[Security Options](#)."

You may need to un-check this box to enable hit highlighting to work with servers that use older versions of dtSearch Web (before version 7.67).

## Check for Updates

Click **Check for Updates** to check for a newer version of the PDF Search Highlighter from the dtSearch Web Site.

# Technical Information

## How the plug-in works

### Overview

The dtSearch PDF Search Highlighter is a plug-in for Adobe Reader that makes it possible for search programs to highlight hits in retrieved documents inside Adobe Reader X. Earlier versions of Adobe Reader had this feature built in, so there was no need for a plug-in.

When a user clicks on a link to a PDF file on a web page, the browser loads Adobe Reader as a plug-in and uses it to display the page.

Adobe Reader versions prior to Adobe Reader X provided a way for search programs to send hit highlighting information so search hits could be highlighted in Adobe Reader, and users could click "Next Hit" and "Previous Hit" buttons in Adobe Reader to find hits quickly.

Adobe Reader X no longer supports this URL syntax. dtSearch Corp. has created the plug-in to restore this very useful Adobe Reader feature. Subject to the included license agreement, a user can download the plug-in at no charge from the dtSearch web site [here](#) and install the plug-on on any number of computers within the user's organization. The plug-in may be used with any compatible product or web site -- not just dtSearch products or web sites that use the dtSearch Engine.

### Installing the plug-in

To install the plug-in, download and run [dtSearchPdfHighlighter.exe](#). dtSearchPdfHighlighter.exe is a self-extracting ZIP archive containing an MSI installation package, dtSearchPdfHighlighter.msi.

The installer adds the following to the user's system:

- A folder in C:\Program Files named "dtSearch\dtSearch PDF Search Highlighter" with dtswebhits.api (the plug-in), dtspdfcfg.exe (a configuration program for the plug-in), and some HTML help files.
- A Start Menu shortcut for the configuration program.
- Copies of dtswebhits.api in the Adobe Reader and Adobe Acrobat plugins folders.
- A registry entry that adds "dtsPdfWh" to the Internet Explorer User Agent string. (For the reason this is necessary, please see "[Internet Explorer Compatibility](#).")
- A Browser Helper Object for Internet Explorer is installed to enable hit highlighting to work in Internet Explorer. (In a "Custom" installation, this can be disabled.)

Uninstalling the plug-in removes all of the above items from the user's system.

The plug-in uses settings stored under this registry key:

```
HKEY_CURRENT_USER\Software\dtSearch Corp.\dtSearch\PlugIns\dtswebhits.api
```

Each time a PDF file opens in Adobe Reader, the plug-in checks the URL used to open the file for hit highlighting data. If the URL includes a reference to highlighting data, the plug-in verifies that the reference to the highlighting data complies with the user's [security settings](#), downloads the highlighting data, and uses it to add hit highlighting to the PDF file in Adobe Reader.

## How hit highlighting data is included in URLs

The format for the URLs that provide hit highlighting information is:

```
http://www.example.com/sample.pdf#xml=http://www.example.com/hits.xml
```

The #xml= portion of the link points to a URL that returns an XML stream describing the location of the hits in the PDF file. The format of the XML file is described in this document:

*Adobe Technical Note 5172 -- Highlight File Format*

<http://partners.adobe.com/public/developer/en/pdf/HighlightFileFormat.pdf>

The #xml= portion of the URL can request the XML from a script or program, like this:

```
http://www.example.com/sample.pdf#xml=http://www.example.com/getPdfHits.asp?idoc=5
```

The dtSearch plug-in supports this URL syntax and also supports a variant, using ?xml= instead of #xml=. The variant is needed because the #xml= URLs are not visible to the plug-in when Adobe Reader is running inside a WebBrowser control.

## Security Options

The plug-in has two categories of security settings: (1) settings that limit which servers can generate hit highlighting information based on the location of the PDF file, and (2) a verification protocol to ensure that the highlighting server is able to generate hit highlighting data.

These are intended to guard against use of the plug-in to exploit security vulnerabilities in web sites that do not sufficiently guard against cross-site scripting and cross-site request forgery attacks.

### Servers that can generate hit highlighting information

The default settings only allow a site to hit-highlight its own PDF files. In cases where a site hit-highlights content located on other web sites, the Security Options dialog box allows the user to enter the web site as a "trusted" site that is allowed to do this.

### Validation of highlighting script

To prevent use of the plug-in to send forged requests to web sites, the plug-in will send a standard validation request to make sure the target URL is really a PDF search highlighter. The validation request replaces the query in the original URL with "IsPdfHighlighter", and expects a response that contains "YesPdfHighlighter".

For example, suppose a user clicks this link:

```
http://www.example.com/harmless.pdf#xml=http://www.example.com/shopping.asp?buy=Product&ShipTo=Attacker
```

Adobe Reader will open the PDF file, and the dtSearch plug-in will see the #xml= in the URL. To verify the target URL, before requesting the highlighting data, the dtSearch plug-in will first send this request:

```
http://www.example.com/shopping.asp?IsPdfHighlighter
```

A shopping script that is not aware of the PDF search highlighting protocol will not respond with the specific "YesPdfHighlighter" code. If the server does not respond with this text, the dtSearch plug-in will decide the server does not support PDF highlighting and the request will not be sent.

To preserve backward compatibility with servers that have not been updated, users will have the option to disable this verification.

## Information for Developers

### Internet Explorer Compatibility

When PDF files are opened inside an Internet Explorer window, the #xml= syntax is not visible to the plug-in, so it cannot access the hit highlighting information. To work around this, the plug-in installs a browser helper object (BHO) in Internet Explorer.

For situations where the browser helper object is not used, such as applications that embed the WebBrowser control, the plug-in supports an alternative syntax, ?xml= instead of #xml=, that works without the browser helper object.

### Changes needed to web sites that use dtSearch Web

If your web site uses dtSearch Web, upgrade to dtSearch Web 7.67 to make the site compatible with the plug-in.

Your users who have Adobe Reader X will need to download and install the plug-in, using this download link:

```
http://download.dtsearch.com/pdfhl/dtSearchPdfHighlighter.exe
```

### Changes needed to dtSearch Publish applications

(1) Upgrade the CD/DVD folder to use version 7.67 of lbview.exe, which is able to highlight hits using the plug-in. To do this, run the dtSearch CD Wizard and click "Change CD Settings" and check the box to "Update program files on the CD".

(2) When the CD starts, users who do not have the plug-in installed will see the page in /builtin/getplugin.html which has a prompt and link to download and install the plug-in.

### Changes needed to existing server-based applications

Server-based applications should support the [IsPdfHighlighter verification protocol](#). To do this, your script that responds to PDF hit highlighting requests should respond to an IsPdfHighlighter request with a response that contains "YesPdfHighlighter". To make troubleshooting easier, you can embed this in a web page so you can see the response in a web browser. Example:

```
<html><head><title>Pdf Highlighter Confirmation</title></head><body><!--  
YesPdfHighlighter --><P>This server supports PDF highlighting</P>
```

## Changes needed to existing desktop applications

Desktop applications can highlight hits in PDF files by embedding a WebBrowser control and navigating to a URL that provides the location of the PDF file and the highlighting data.

Desktop applications should continue to operate as before if Adobe Reader 9 or earlier is installed.

If Adobe Reader X is installed, two changes are needed:

(1) The URL your application uses has to use ?xml= instead of #xml=. This is necessary because the #xml= syntax does not work in an embedded WebBrowser control.

(2) LbvProt.dll does not work with Adobe Reader X in protected mode (in protected mode, Adobe Reader runs in a separate process from the browser), so your application should instead generate file-based URLs to send highlighting data to the plug-in:

- Call SearchResults.MakePdfWebHighlightFile as before to generate the highlight data.
- Write the highlight data to a temp file.
- Generate a URL with the location of the PDF file and the temp file like this:  
c:\documents\example.pdf?xml=c:\temp\highlight.xml
- Call the WebBrowser's Navigate method with the URL to open the PDF file.

## Distribution of the plug-in

Subject to the included license agreement, a user can download the plug-in from the dtSearch web site and install the plug-in on any number of computers within the user's organization. The plug-in may be used with any compatible product or web site -- not just dtSearch products or web sites that use the dtSearch Engine.

To comply with Adobe licensing requirements, the plug-in must be deployed using the dtSearchPdfHighlighter.msi installer and must be distributed only from the dtSearch download URL (<http://download.dtsearch.com/pdfhl/dtSearchPdfHighlighter.exe>).

Adobe Reader plug-ins can only be created using a digital signature issued by Adobe, which dtSearch Corp. must renew each year in January with permission from Adobe. Each time the license is renewed, the plug-in must be updated with a new digital signature, and renewal is subject to Adobe's discretion.