

Using LEADTOOLS PDF File Features to Enhance Google Drive Search

Introduction

Cloud services such as Google Drive continue to grow in popularity each year as a safe, secure and convenient way to store and back up your documents, images, music and other files. For users with a large amount of data in the cloud, searching and finding your files again can become problematic. Most search features are limited in their scope, and only take advantage of the file name or, for file formats such as PDF, the text within the file itself. Therefore some level of customization or enhancement may be necessary to take full advantage of your Google Drive cloud storage.

Searching for a PDF may be easier than searching for an MP3 or JPEG, but Google Drive has some limitations with the format as well. For example, let's say you scan an invoice or bank statement and save it as a PDF. Even if you have a scanner or software that extracts the text with OCR, you still might not have a reliable way of searching for that document. The text would likely have words for the name of the company and date, but it might be lacking keywords that you find useful for archiving and finding the document later such as "bank", "insurance", "paid with PayPal" and so on.

This is exactly the kind of information you would want to include in the Keywords metadata of your PDF file when saving it, but Google Drive doesn't use this metadata in its search index. Therefore you can use the LEADTOOLS PDF SDK to read and edit the file metadata, and then update the file's [IndexableTextData](#) property in Google Drive. In the white paper that follows, we will show how to read and write the PDF keywords metadata, update the file on Google Drive, interface with your local Google Drive database, and do all of this within a single right-click context menu in Windows Explorer.

Creating the Right-Click Context Menu

When using a service such as Google Drive which comes with a desktop application to automatically sync files on your computer with your online cloud drive, a full blown application isn't necessary. A more practical approach is to add a context menu item that appears when you right-click on a PDF file. After the command is added to the registry, you can right-click on any PDF file and select "Update File Keywords," which will pass the file name as an argument to the application.

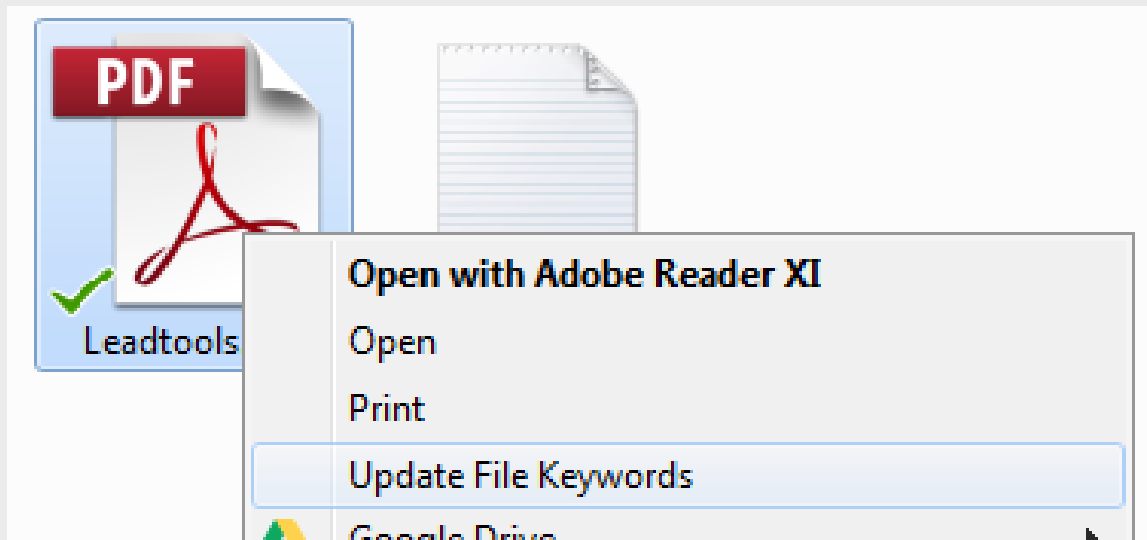
```

using (RegistryKey pdfTypeRegKey =
    Registry.ClassesRoot.OpenSubKey(".pdf"))
{
    // Create path to registry location
    string regPath = string.Format(@"{0}\shell\{1}",
        (String)pdfTypeRegKey.GetValue(null), "UpdateFileKeywords");

    // Add context menu to the registry
    using (RegistryKey key = Registry.ClassesRoot.CreateSubKey(regPath))
    {
        key.SetValue(null, "Update File Keywords");
    }

    // Add command that is invoked to the registry
    string menuCommand = string.Format("\{0}\\" \"%L\"",
        Application.ExecutablePath);
    using (RegistryKey key = Registry.ClassesRoot.CreateSubKey(
        string.Format(@"{0}\command", regPath)))
    {
        key.SetValue(null, menuCommand);
    }
}

```



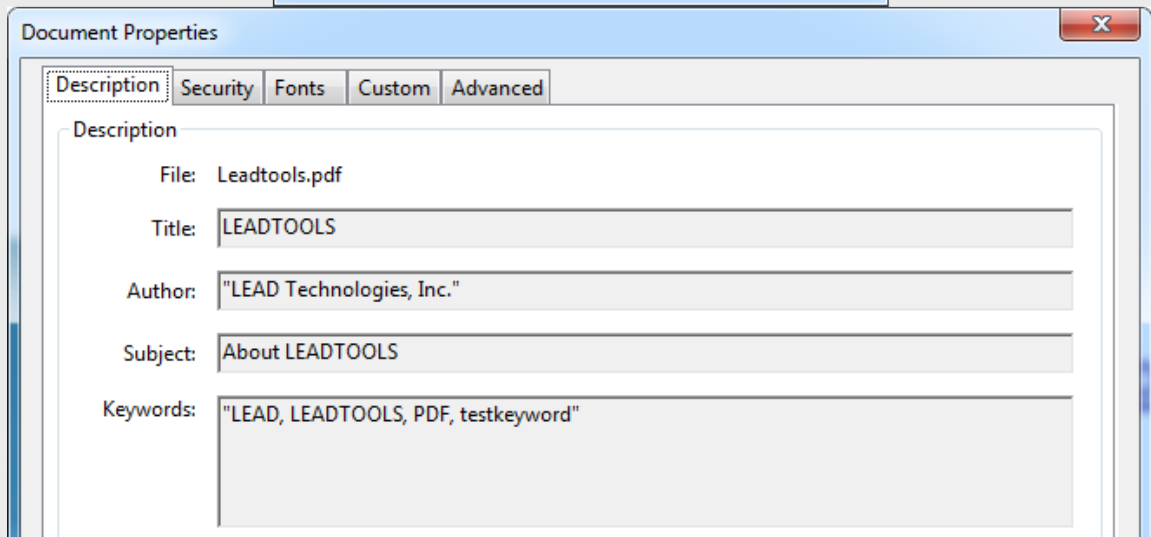
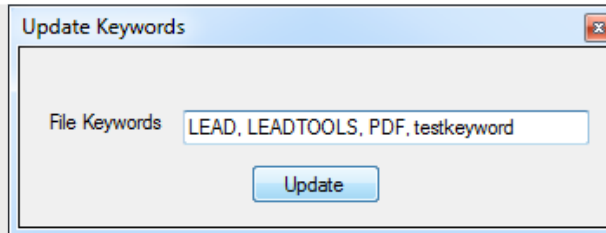
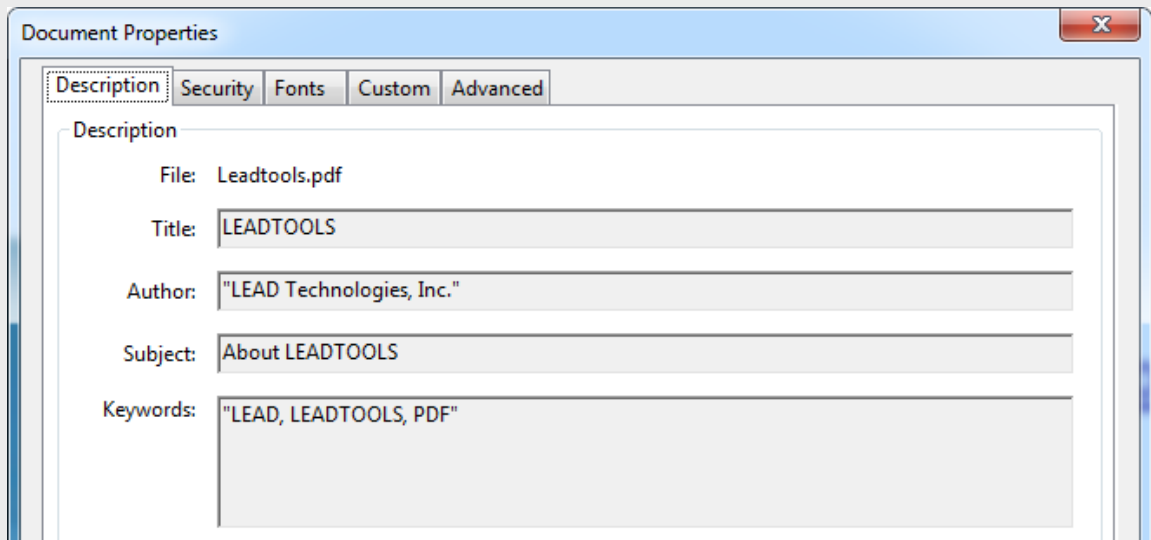
Using LEADTOOLS to Update PDF File Keywords Metadata

Now that the foundation for our application is laid, we must update the keywords within the PDF File. LEADTOOLS includes comprehensive PDF reading, writing and editing capabilities in a programmer-friendly SDK which allows for the direct modification of PDF file properties, searchable text, bookmarks and more. When our application loads from the right-click menu shell command, it will use the LEADTOOLS `PDFFile` object to retrieve the keywords and display them in the textbox for editing.

```
PDFFile _document = new PDFFile(fileName, password);  
_document.Load();  
_txtKeywords.Text = _document.DocumentProperties.Keywords;
```

Saving is just as simple, requiring only a few lines of code. As you can see, the document properties of the PDF are correctly updated with the new keywords.

```
_document.DocumentProperties.Keywords = _txtKeywords.Text;  
_document.SetDocumentProperties(fileName);
```



Updating Google Drive

Finally, a few more steps must be taken in order to wrap up our enhancement to Google Drive's PDF search. The keywords and other metadata properties within PDF files are useful and powerful features, but Google Drive does not use them within its search algorithm. However, each file in Google Drive has the `IndexableTextData` property which can be modified when using the Google Drive API.

When using the Google Drive desktop sync application for Windows, it uses a local SQL database to keep track of the local files and their online information. In order to complete this operation we must get the `fileId` that matches the local file we just updated. Depending on how your Google Drive folder is organized, you may need additional queries to recursively find the file within subfolders. However, once you acquire the `inode_number` that matches the PDF file name you passed through the right-click menu command, you can get the `fileId` from the database and call the Google Drive web service.

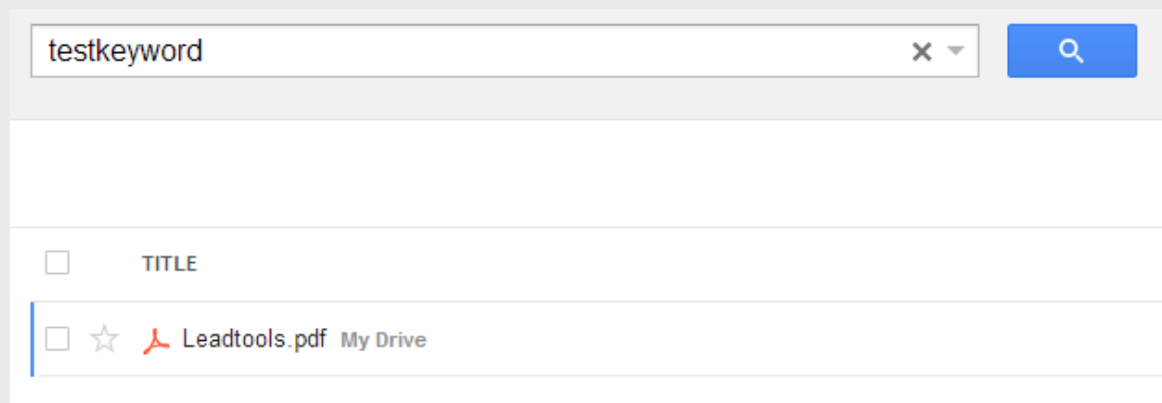
```
// Get resource_id for the target file (formatted "type:resource_id")
sqlitecmd.CommandText = "SELECT resource_id FROM mapping where
    inode_number='" + fileInodeNumber + "'";

reader = sqlitedb.ExecuteReader();
reader.Read();

String fileId = reader["resource_id"].ToString().Split(':')[1];
reader.Close();

File file = googleDriveHelper.GetFile(fileId);
file.IndexableText = new File.IndexableTextData();
file.IndexableText.Text = _document.DocumentProperties.Keywords;
googleDriveHelper.UpdateFileMetadata(file);
```

Now you can search your Google Drive for your custom PDF keywords, increasing the already incredible value of Google Drive's free cloud storage service.



Conclusion

This is just one of many real world solutions you can tackle with LEADTOOLS. Its state of the PDF SDK is the most flexible and powerful product in its class, and LEADTOOLS offers an incredible value with its comprehensive family of toolkits for raster, document, medical and multimedia imaging. For more information on how LEAD Technologies can image-enable your application and boost your ROI, visit www.leadtools.com to download a free evaluation, or give us a call at +1-704-332-5532.

SALES: (704) 332-5532
SALES@LEADTOOLS.COM

SUPPORT: (704) 372-9681
SUPPORT@LEADTOOLS.COM



LEAD TECHNOLOGIES, INC.
1927 SOUTH TRYON STREET
SUITE 200
CHARLOTTE, NC 28203

About LEAD Technologies

With a rich history of over twenty years, LEAD has established itself as the world's leading provider of software development toolkits for document, medical, multimedia, raster and vector imaging. LEAD's flagship product, LEADTOOLS, holds the top position in every major country throughout the world and boasts a healthy, diverse customer base and strong list of corporate partners including some of the largest and most influential organizations from around the globe.

LEADTOOLS[®]
THE WORLD LEADER IN IMAGING SDKs

