How to Create a Searchable Library of PDFs

Using INM Impressario 3.0 and INM VizionDB 4.1

For Macromedia® Director®

Contents

Copyright and trademark notice

Introduction

Allowing users to search through a collection of PDF resources has always been a popular feature in interactive publications. Previous versions of INM Impressario had a tool called the PDF Indexer, which combined the power of the INM V12 Database Xtra and INM Impressario to enable this functionality.

We have now improved this functionality by combining INM VizionDB and INM Impressario 3. The PDF indexing functionality, now called "**Building a Searchable Library**" is done at authoring time directly within the INM VizionDB Editor.

When implementing your Searchable Library in Director, Impressario 3 now includes a Flash-based **Search Panel**, similar to the Adobe search panel, which you can drag and drop on stage with no coding necessary.

This search panel incorporates the new and more powerful search features provided by INM VizionDB, including the ability to do Boolean and wildcard searches, specify proximity of two or more search terms, and display relevancy scores in your results list.

This manual contains all the information you need to build a Searchable Library and integrate the Flash-based search panel in Adobe Director. It also provides a list of relevant methods of INM VizionDB and INM Impressario that you can use to customize the Searchable Library for your own project needs. Further details for the scripting API's of INM VizionDB and INM Impressario are provided within those resources.

Requirements

- Director MX or higher
- PDF documents that conform to PDF Standard 1.4 - 1.6
- INM Impressario 3.0+ for Director *Full Version*
- INM VizionDB Desktop version 4.1+
- Mac OS X 10.2.8 - Mac OS X 10.4
- Windows NT, 2000, XP

Using the trial version

In order to deliver the Searchable Library feature in your project, you need to purchase a license for both the Full version of INM Impressario 3, and INM VizionDB Desktop 4.1. However, you can try out this feature using the trial versions of these two products. The trial versions provide all the functionality of the licensed versions, except that there are splash screens that appear to warn that you are using a trial version, and you eventually need to purchase a real license.

Special Package Pricing

Due to the popularity of the Searchable Library solution, you can purchase licenses for both INM Impressario and INM VizionDB together, at a significant cost savings. See **The Essential Pack for Director** on our web site for more details:

www.INM.com/products/essential

Installing the required files

1. Download the installers for INM VizionDB 4.1 and INM Impressario 3. When you download the installers, a Trial Key for each one will be sent to your e-mail address. Keep these trial keys handy so that you can enter the appropriate key for each installer when prompted.

2. First run the installer for INM VizionDB 4.1. If you are on Mac, also run the separate installer for the INM VizionDB Editor.

3. Then run the installer for INM Impressario 3. This will place the necessary PDF Libraries in the correct folder location to be accessed by the INM VizionDB Editor.

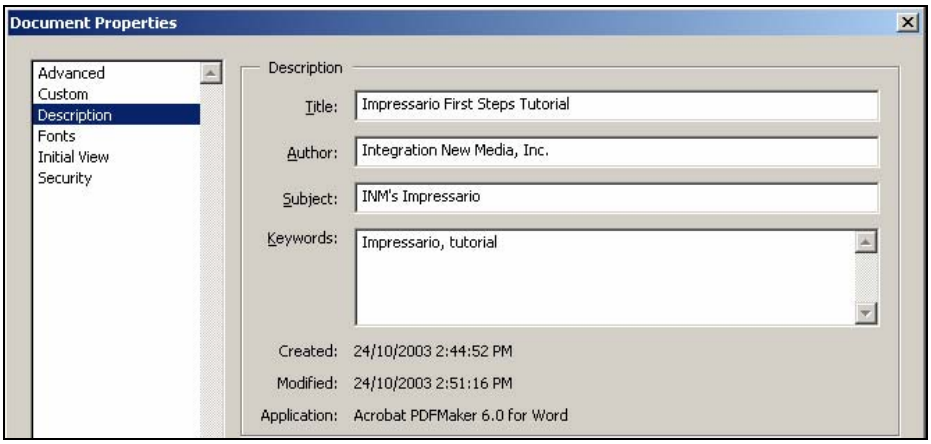
Once you purchase the full licenses for the two products, you will enter your username, company name and license keys in the **Enter Key** dialogs for each product from the Director **Xtras** menu.

Adding metadata to your PDFs

The Searchable Library Wizard of the INM VizionDB Editor not only indexes the full text of your PDF documents, but can optionally create database fields for the metadata (document properties) that describe your PDF documents, such as Title, Author, Subject, Keywords, Date Created and Date Modified, as well as any custom properties you have defined..

If you want to allow your end-users to be able to search for content based on metadata, you need to save the documents with the appropriate information, to classify them in a way that makes sense for your application.

In Acrobat 6 or higher, select **File > Document Properties...** from the menu bar and view the **Description** pane (in Acrobat 5.0, select **File > Document Properties > Summary...**), or type **Ctrl-D** on the keyboard. The following window appears where you can add or change the current document's properties.



Document Properties window in Acrobat 6.0

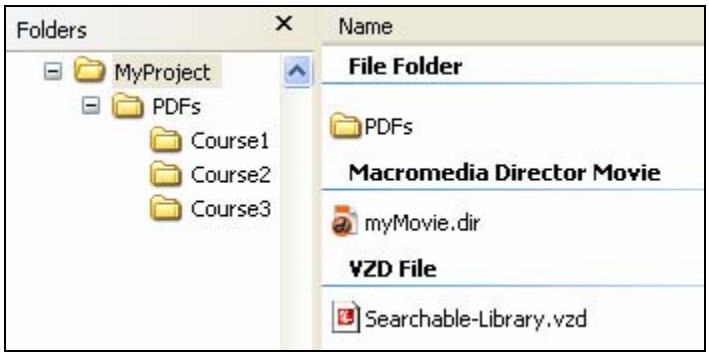
Note that the "Created" and "Modified" dates are controlled by Acrobat and cannot be changed from this window.

Organization of the PDF files

Gather all the PDFs that you want to include in the Searchable Library within a folder hierarchy, where the top-level folder (the root) will contain the INM VizionDB database that contains the Searchable Library. The Searchable Library function can index PDF files that are located either in the same folder, or in subfolders of the open INM VizionDB database.

You will also need to place your Director movie at this root level when deploying your project because the Searchable Library database stores the PDF documents' paths **Relative** to `the moviePath`.

In the example folder structure, shown below, the database named "Searchable-Library.vzd", as well as the movie named "myMovie.dir" are located the top level of the folder hierarchy and the PDFs are located within subfolders of the "PDFs" folder.



Typical folder structure for a Searchable Library project

How to build a Searchable Library of PDFs

Even if you don't have a collection of PDFs to use, you can test out the functionality of the Searchable Library using the sample PDFs that come with the Searchable Library sample movies. If you already have your own PDFs to work with, feel free to follow the steps below using those.

Start by opening the INM VizionDB Editor:

- On Windows click **Start > All Programs > Integration New Media > INM VizionDB > INM VizionDB Editor**.
- On Mac OS X open the INM VizionDB Editor from **Applications > Integration New Media > INM VizionDB Editor**.

Create an INM VizionDB database

1. From the INM VizionDB Editor menu, click **New Database > From Scratch...**
   (If you already have an INM VizionDB database in which to create your Searchable Library, just open that existing database)

2. Choose a language for your database.

3. Browse to the root folder where your PDFs are located (see Organization of the PDF files), give the new database a name and click **OK**.

4. The **Modify Database Structure** dialog opens by default. Click **Create Searchable Library...** from that dialog. The Searchable Library Wizard opens.

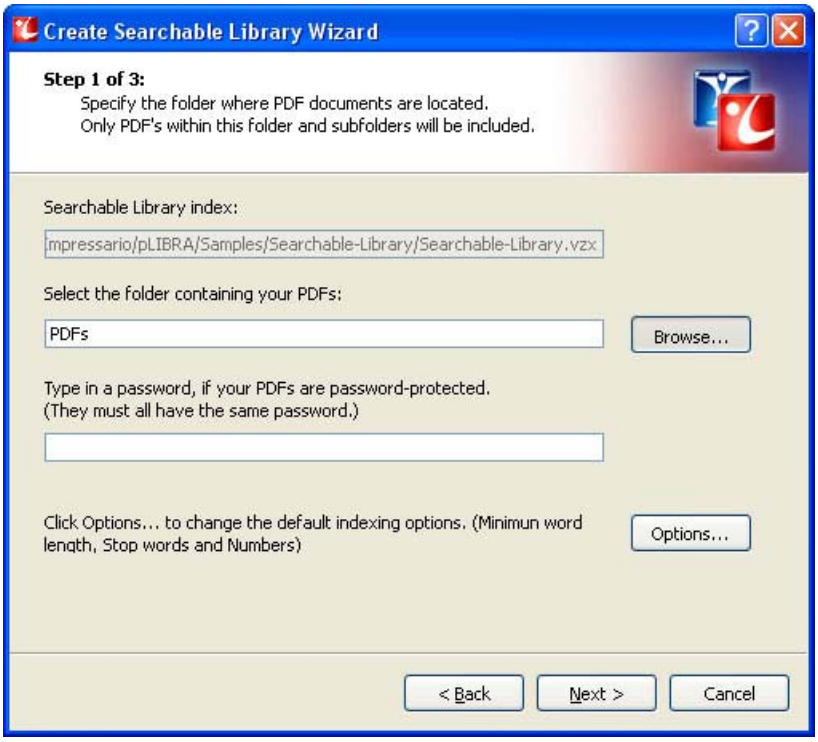
5. Proceed to the next section: The Searchable Library Wizard

The Searchable Library Wizard

If you followed the steps above to create a new Searchable Library from scratch, you have already opened the Searchable Library Wizard. If you are starting with a previously existing INM VizionDB database, click **Tools > Create Searchable Library** to start the wizard.

Step 1. Select the folder to Index

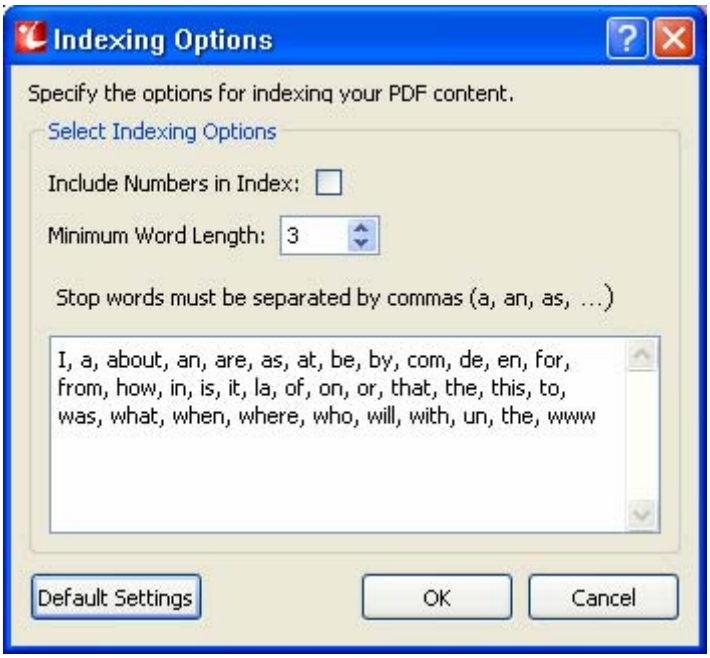
The first step is to select the root folder name at which the tool will begin reading and indexing the PDFs. The Searchable Library wizard will only store the PDF document pathnames relative to the database location; that is why the database should also be created at this root level.



Step 1. Specify the folder where the PDFs are located.

1b. Indexing options

Click the Options button to change any of the default indexing options. Indexing options include the ability to omit certain common terms from the index, called **Stop Words**, and to specify a minimum length word to be indexed. The default Stop Words depend on the language chosen for the database. However, you may modify this list and your selections are saved for the current database.

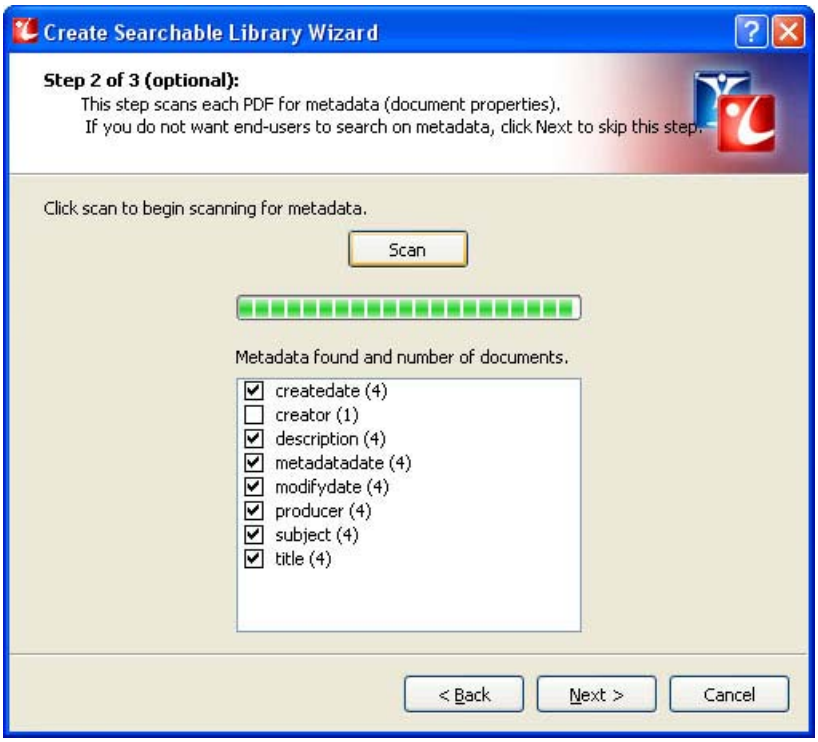


Indexing Options dialog

Step 2. Scan for Metadata

Before indexing the text within the PDFs there is an initial **Scanning for Metadata** process which retrieves the document metadata (properties such as Title, Author, Subject, CreationDate, etc., including custom properties) from each PDF. Once this process is complete, you are presented with a list of the metadata found and the number of PDFs within the set that contain each property. You then have the choice to include or exclude each field of metadata from the SEARCHABLE_TABLE structure, by checking or unchecking the box next to that metadata.

You must complete this process if you want to allow users to search based on metadata or if you want to display PDF titles, rather than just the filenames, in the search results list that your end-user sees. Otherwise, if you just want to allow users to search the complete PDF text and you are content to display the relative path names, you may skip the scanning process.



Step 2. Scan for Metadata

Step 3. Build the Index

In this step, the text of each PDF document is read and indexed so that fast searches can be performed by the end-user. The amount of time it takes to index the PDFs depends directly on the amount of text in your PDFs.

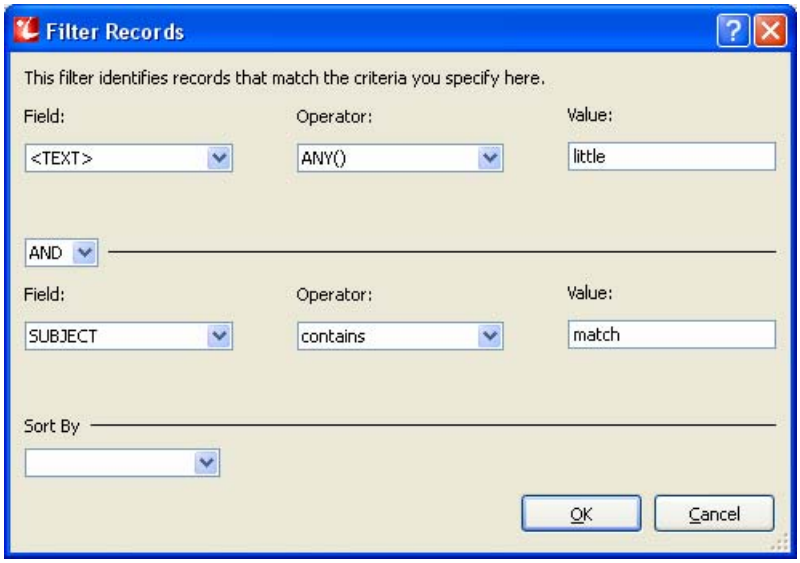
When this step is done, a new index file is created next to your INM VizionDB database. This file must be kept next to your INM VizionDB database in order to perform any searches on the indexed PDF content.

After indexing all PDFs, the conclusion screen displays the total number of PDFs and the total words indexed.

Validate the Results

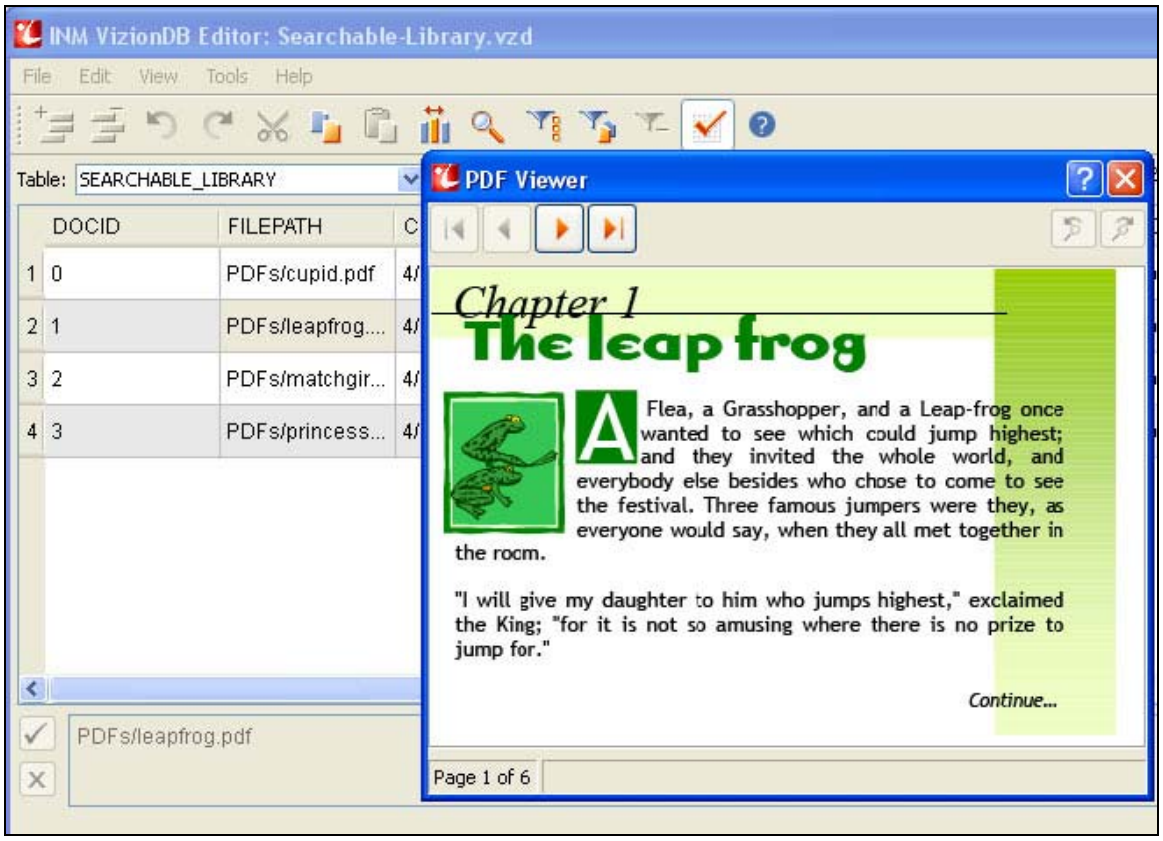
Once you are done building the index, you can view the contents of the Searchable_Library table within the INM VizionDB Editor's main screen. You can sort the PDFs by any of the metadata fields (filename, title, etc.).

To search within the VizionDB Editor and validate search results using the **View > Filter with Criteria...** function, or using the **Query Designer Tool.**



Filter with Criteria allows you to do quick spot-checking of your PDF content

Double-click a PDF record in the main table to open a PDF viewer from which you can flip through the pages to verify the document's contents.



Double-click anywhere within a PDF record in the main table to view the document on screen

Click **Tools > Searchable Library Info...** to view the indexing options that were in effect when the Searchable Library was created, as well as the total number of PDFs, total number of words indexed, and a list of the indexed words, which you can export to a text file.

For more information on using the INM VizionDB Editor, consult the INM VizionDB Editor User Manual.

Implementing a Search Panel in Your Project

Now that you have a Searchable Library VizionDB database, you need to create an interface that end-users interact with to search and retrieve PDF content. The easiest way to do this is to use the pre-defined Flash-based **Search Panel** that comes with Impressario.

Steps to integrate the Search Panel into your interface

Follow these steps to implement a Searchable Library in your Director movie.

Step 1. Insert an Impressario member

Create a Director movie and insert a member of type Impressario, if you have not already done so. (See instructions for **Inserting an Impressario Member** in **Xtras > INM Impressario > Help Online** files if you need more information.)

You may choose to link a PDF document to your member, to display an initial PDF, or leave the filename property empty initially. The PDF documents that the user selects from the Search Results will be linked and displayed dynamically at runtime.

Step 2. Insert INM VizionDB Database and Recordset members

Next, you need to insert your INM VizionDB cast members.

Click **Insert > Integration New Media > INM VizionDB Database**. Browse to locate the INM VizionDB database containing the Searchable Library. Click **OK** to close the INM VizionDB Database member dialog.

An INM VizionDB Recordset member is automatically created for you in your Director cast. Give both the INM VizionDB Database and Recordset members names "DB" and "RS" so that you can easily refer to them later.

Double-click the Recordset member to open its dialog. Make sure the **Table** selected in the dialog is "Searchable_Table" and that all fields are selected.

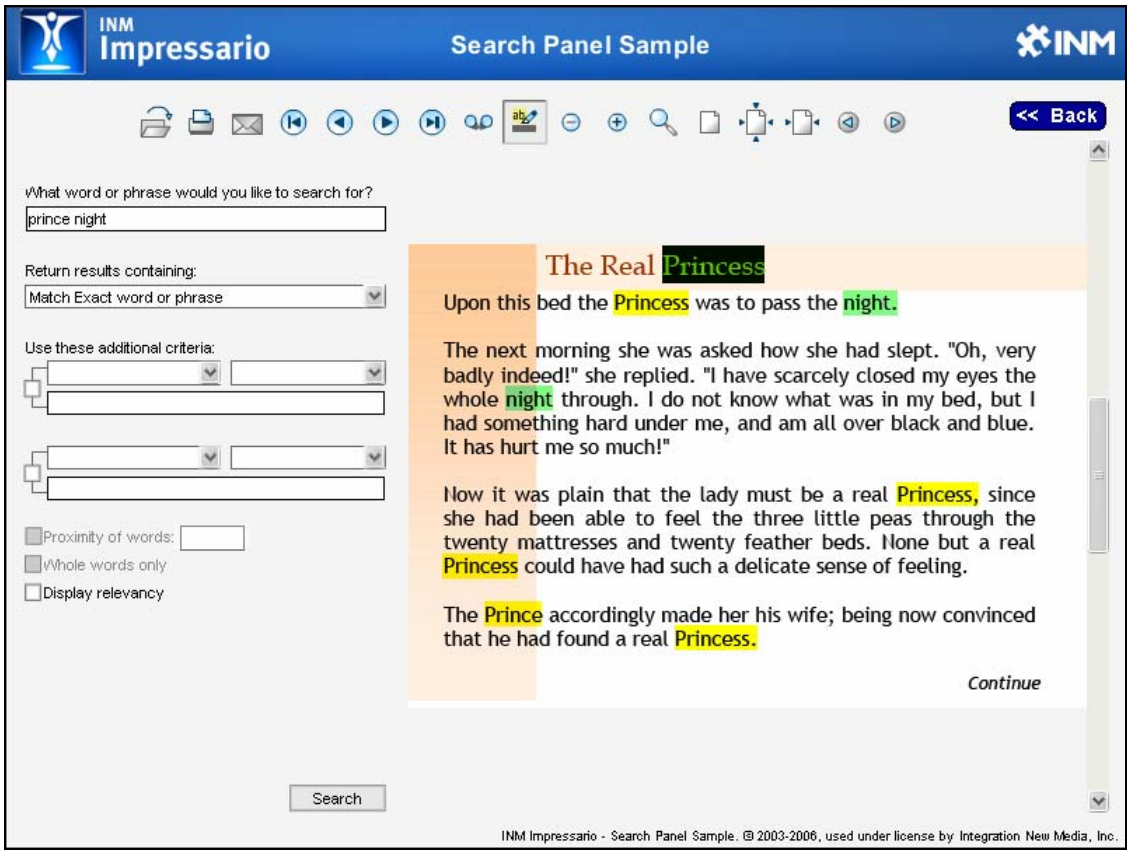
Step 3. Insert the Search Panel

Once you have your INM Impressario sprite on stage and your INM VizionDB members configured, you will need to open the castlib named Searchable Library.cst that was included with the Searchable Library Toolkit. If you

copied this file to the Libs/INM Impressario Tools folder within the Director folders, you can access it through the Library Palette.

The castlib contains a Flash member named **Impressario Search Panel**, and a corresponding behavior named **Search Panel Behavior**.

Drag the **Impressario Search Panel** member to the stage. Position the panel on screen so that it is not overlapping the Impressario sprite.



Search Panel sample movie

Step 4. Attach the Search Panel Behavior

Drag the **Search Panel Behavior** member from the Searchable Library castlib onto the Search Panel sprite on stage. The Search Panel Behavior parameters dialog opens.

Make sure your Impressario sprite channel is correctly selected in the **Impressario sprite channel** dropdown and click OK to close the behavior dialog.

Step 5. Run your movie

Run your movie and test out the options of the Search Panel:

Type one or more search terms and choose the type of search you want:

- **Match Exact word or phrase** finds only PDFs that contain the exact sequence of words entered.
- **Match Any** returns the PDFs that contain any of the words entered.
- **Match All** returns only PDFs that contain all of the words entered.
- **Complex search:** allows you to enter complex search criteria. It interprets Boolean operators, such as &, | and ! (AND, OR and NOT) as well as wildcard characters (* and ?).

Click **Search** to retrieve the list of matching documents.

Click one of the PDF documents returned in the **Search Results** list.

The PDF appears on screen, with the search terms highlighted in different colors.

Try a search on two or more terms with **proximity** selected. Enter a number that indicates the maximum number of words that can separate each of the search terms within the PDF contents and see how the results differ.

Creating Custom Search Interfaces

If you prefer to implement your own search interface in Director, rather than use the Search Panel that comes with Impressario, you can use the Searchable Library behaviors provided, or even code your own search functionality directly, using the scripting API for INM VizionDB and INM Impressario. The sections below describe each of these options.

Searchable Library Behaviors

For a customized look and feel, without needing to code in Lingo, try the Searchable Library behaviors, which are accessible via Director's Library Palette. An example movie, named **Search-Sample.dir** is provided in the Searchable Library package, so that you can get an idea of how the behaviors work.

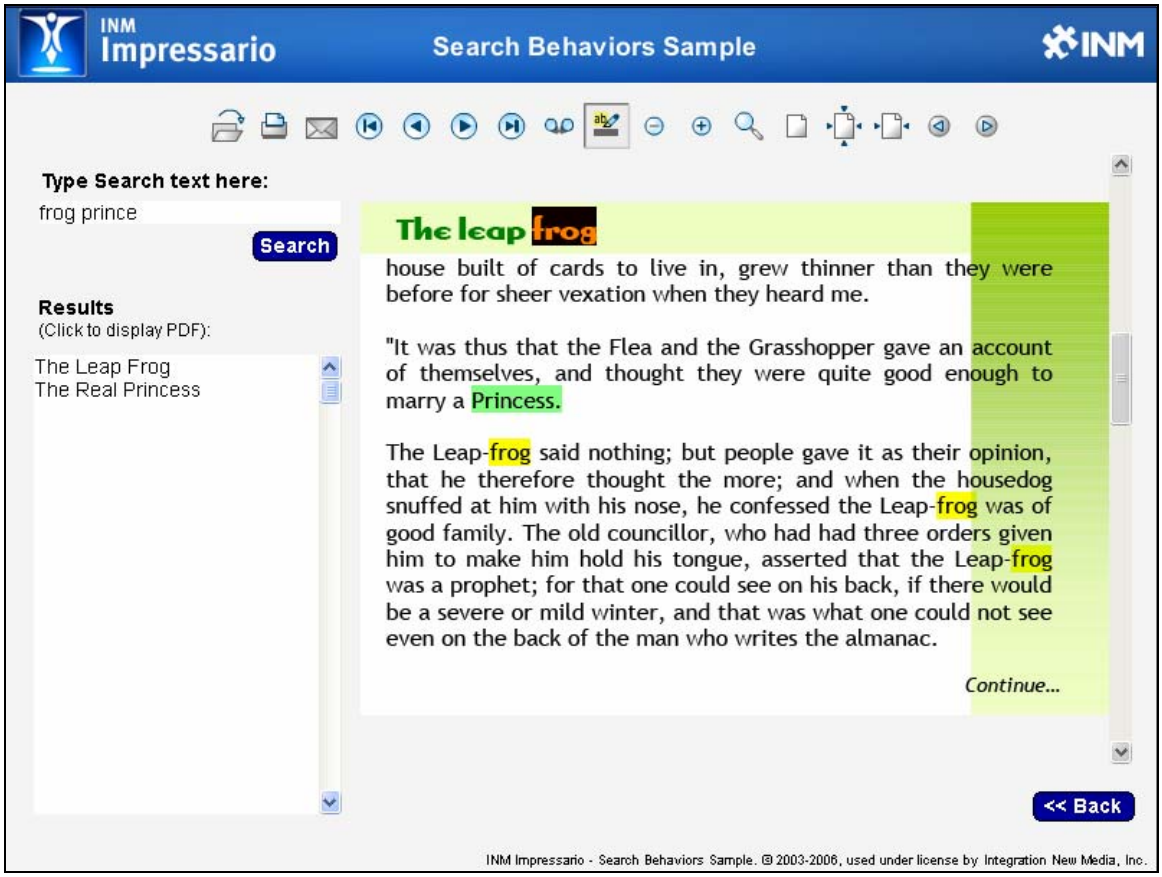
Step 1. Create search fields and buttons

To implement the search functionality of our sample movie in your own project, you will need:

- One editable field or text member for the user to enter the text to search for within the PDF
- One button that performs a search within the text of each PDF document
- One non-editable field to display the list of PDF documents that match the search criteria
- An Impressario Sprite, to display the PDF document that the user selects from the results field

Optional fields and buttons:

- One editable field or text member for the user to enter the text to search for within a document property
- Two editable fields to input the beginning and end dates for a search by date range
- At least one button to search within a particular PDF document property
- One button to search for documents either created or modified within a specified date range



Search Sample with Behaviors

Step 2. Insert the Impressario and INM VizionDB members

The Searchable Library behaviors can only be used if both INM Impressario and INM VizionDB are installed and you have members of type INM Impressario, INM VizionDB Database and INM VizionDB Recordset in your movie.

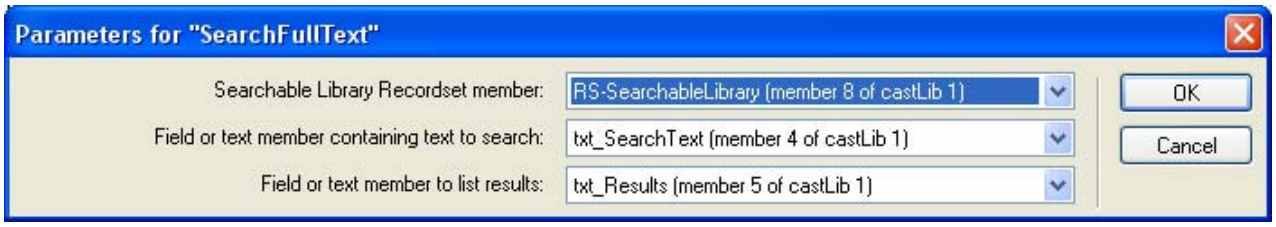
If you have not already inserted INM Impressario and INM VizionDB members into your movie, follow the first two steps of the section entitled: Steps to integrate the Search Panel into your interface.

Step 3. Use the Searchable Library behaviors

From Director's menu bar, click **Windows > Library Palette**, to open Director's library palette. Locate the **INM Searchable Library Behaviors** by clicking the menu button in the top left corner of the palette. There are just four behaviors:

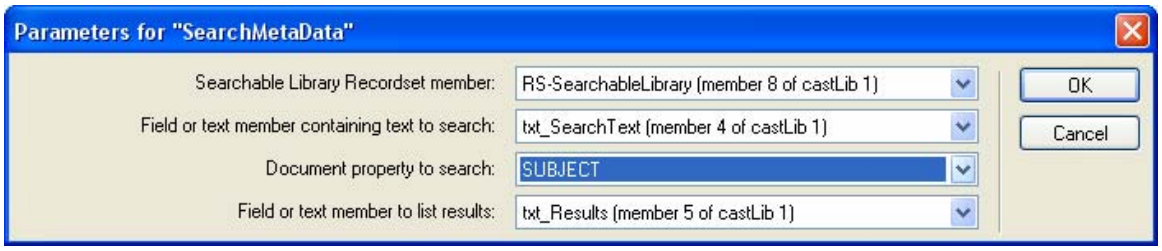
- **SearchFullText**
- **SearchMetaData**
- **SearchDateRange**
- **OpenFoundPDF**

1 Drag the behavior named **SearchFullText** onto the button that will search through the text of the PDF documents. The parameters dialog box opens and you must specify:
   - which INM VizionDB Recordset contains the Searchable Library
   - which member contains the text to search for, and
   - which member to list the results in



The search function used in the **SearchFullText** behavior is Search_ANY; it will retrieve PDFs that contain **any** of the search terms entered.

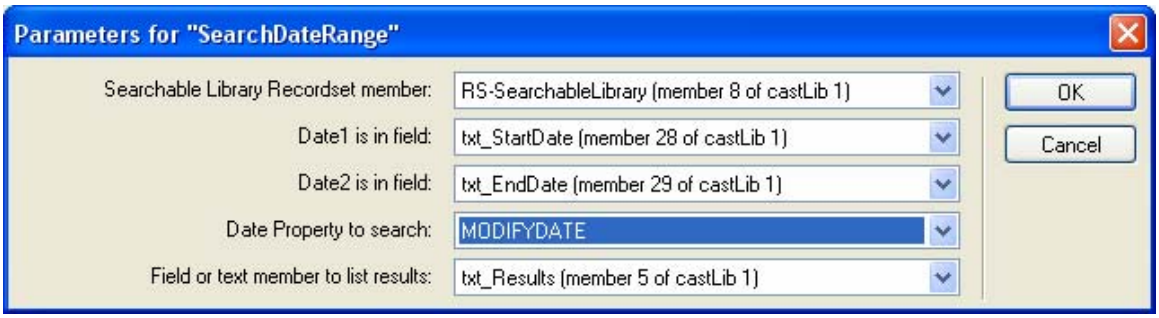
2 Drag the behavior named **SearchMetaData** onto a button to search for PDFs based on a document property, such as Title, Author, Subject, Keyword, etc. In the parameters dialog box, specify:
   - which VizionDB Recordset contains the Searchable Library
   - which member contains the text to search for,
   - which PDF document property to search
   - which member to list the results in, and



Note: To use the **SearchMetaData** behavior, your PDF documents must be saved with the appropriate document property information.

3 Drag the behavior named **SearchDateRange** onto the button that will search for documents either created or modified within the date range specified. In the parameters dialog box, specify:

- which VizionDB Recordset contains the Searchable Library
- which field contains Date 1 (the beginning of the range)
- which field contains Date 2 (the end of the range)
- which field to put the results in, and
- which date property to search



4 Drag the behavior named **Open Found PDF** onto the Director field or text member that will contain the results of your search. This behavior requires you to identify the Searchable Library Recordset, the Impressario sprite channel in which to open the selected document, the member containing the text to find within the document and a checkbox to display the Find dialog upon opening the document.

When you select a PDF document listed in the search results list, the document is opened and the search text is highlighted if it is found within the text of the document. (If you are doing a search by document property, the search text may not be found within the document's text).

Scripting API for the Searchable Library

If you are comfortable using Lingo and SQL, and you want the flexibility to define your own search interface or dynamically conduct a search based on certain user interactions through your application, you can use the INM VizionDB and INM Impressario methods and properties outlined below.

INM VizionDB SQLSelect Search Operators

The following Search operators are used within a SELECT query passed to the SQLSelect() method. They are only recognized by INM VizionDB and can only be used with the Searchable_Library table. Not that the search terms must be enclosed by single quotes inside the SQLSELECT string.

SEARCH_EXACT: INM VizionDB will return the PDF documents that match the exact phrase entered (but ignoring punctuation).

Examples:

```
-- simple search for exact text entered by the user
strSQL = "SELECT FROM SEARCHABLE_LIBRARY WHERE WHERE " & \
"SEARCH_EXACT('" & member("user-entry").text & "')"
member("RS-SearchableLib").SQLSelect(strSQL)
```

SEARCH_ANY: Searches for any of the individual words or phrases entered by the user — performs a Boolean OR between words or phrases. Quotes can be used to group words together into a phrase.

Examples:

```
-- simple search for any of the search terms entered by the user
strSQL = "SELECT FROM SEARCHABLE_LIBRARY WHERE " & \
"SEARCH_ANY('dog cat')"
member("RS-SearchableLib").SQLSelect(strSQL)
```

SEARCH_ALL: This function only retrieves documents that contain all of the words specified — the equivalent of a Boolean AND operator between words.

```
-- simple search for documents containing all of the search terms
strSQL = "SELECT FROM SEARCHABLE_LIBRARY WHERE " & \
   "SEARCH_ALL('" & member("user-entry").text & "')"
member("RS-SearchableLib").SQLSelect(strSQL)
```

SEARCH_COMPLEX: This function interprets Boolean operators, such as &, | and ! (AND, OR and NOT) and wildcard characters (***** : any number of unspecified characters and **?** : a single unspecified character). Parentheses are also interpreted to allow the construction of complex search criteria when Search_Complex() is used.

Wildcard example:

```
-- find all PDFs containing words that start with t and end with r
strSQL = "SELECT FROM SEARCHABLE_LIBRARY WHERE " & \
   "SEARCH_Complex('t*r')"
member("RS-SearchableLib").SQLSelect(strSQL)
```

This search retrieves all documents containing words such as "theater", "terminator" and "tear".

Boolean search example:

```
-- find PDFs containing "student" and either "boy" or "girl"
strSQL = "SELECT FROM SEARCHABLE_LIBRARY WHERE " & \
   "SEARCH_Complex('student & (boy | girl)')"
member("RS-SearchableLib").SQLSelect(strSQL)
```

SEARCH_COMPLEX also allows you to specify the proximity of search terms using the following syntax:

Search_Complex(*word1* w:3 *word2* w:5 *word3*')

Proximity example:

```
-- find PDFs containing "accounting" within 5 words of "news"
strSQL = "SELECT FROM SEARCHABLE_LIBRARY WHERE " & \
   "SEARCH_Complex('accounting w:5 news')"
member("RS-SearchableLib").SQLSelect(strSQL)
```

Relevancy ranking

Relevancy ranking is an optional feature that can be applied to any of the Searchable Library functions. It assigns a percentage to each PDF document retrieved based on how relevant that document is compared to the other PDFs that are returned from the search operation. In other words, the relevancy ranking of a particular PDF with respect to the search query will be different, depending on the other PDF documents indexed within the Searchable Library.

Factors that influence the relevancy rank calculation are:

- The number of times the search term occurs within the PDF.

- The number of times the search term occurs across the collection of PDFs.
- The number of words within the PDF.
- The frequencies of words within the PDF.
- The total number of PDFs in the Searchable Library.

When Relevancy ranking is applied to a search query, PDFs that have a relevancy rank of less than 40% are automatically excluded from the results list.

Relevancy ranking can be applied to any of the Searchable Library functions by adding a list of keywords as a second parameter and specifying the field name "Relevancy" in your SQL SELECT query. The "Relevancy" field is a virtual field that is appended to the results.

You can sort your results in order of relevancy, by specifying ORDER BY RELEVANCY in the SQL statement.

Examples:

```
mySQL = "SELECT *, Relevancy " & \
"FROM searchable_library " & \
"WHERE SEARCH_ALL('rabbit flower', 'rabbit flower') ORDER BY Relevancy"

mySQL = "SELECT *, Relevancy " & \
"FROM searchable_library " & \
"WHERE SEARCH_Exact('healthcare provider', 'healthcare plan') ORDER BY
   Relevancy"

mySQL = "SELECT *, Relevancy " & \
"FROM searchable_library " & \
"WHERE Subject LIKE 'Fairytale' AND " & \
"SEARCH_EXACT('happliy ever after') AND " & \
"SEARCH_ALL('prince', 'prince charming castle') ORDER BY Relevancy"
```

Other related INM Impressario methods and properties

Once you have a retrieved a list of PDFs using the INM VizionDB search methods, use the following methods and properties to open the selected PDF document in an Impressario member and perform the Find and highlighting operations.

- **PathType property:** Set the Impressario member's PathType to #RelativeToMovie before setting the #Filename property, because the filename provided by the INM VizionDB database is relative to the movie location.
- **Filename property:** Set the Impressario member's filename to the filename of the PDF selected by the user.

- **Highlighting property:** This is a sprite property. Set the Highlighting of the Impressario sprite channel to True so that when you perform the Find operation, all found words are highlighted.
- **Find method:** This is the sprite method that searches through the PDF document open within the sprite for the first occurrence of the search text specified. The Find method can accept a single search term or phrase (one highlight color is used), or a list of multiple search terms or phrases (a different color is used to highlight each item in the list).
- **SetHighlightColors method:** This is a sprite method that allows you to change the list of RGB colors used to highlight each separate word or phrase of the search text found within the document.

Consult the INM Impressario User Manual for full descriptions of these properties and methods.

Deploying your project

Ensure VizionDB and Impressario members are licensed

When you are ready to deploy your project you need to make sure you have purchased licenses for INM VizionDB 4 and INM Impressario 3. Be sure to enter the real license key and replace the trial key in the Enter Key dialogs, accessed via **Xtras > INM Impressario > Enter Key...** and **Xtras > INM VizionDB > Enter Key...** in Director.

Once you have entered your full license key, open your project, double-click one of your Impressario members in the cast to open the Impressario member options window and click **OK** to make sure the member is stamped as licensed.

Do the same for one of your INM VizionDB Recordset members to ensure that the database is opened and closed at least once after your license is entered.

Files to include

When you publish your projector, you don't need to have the Xtras included within the projector. You will, instead include the following folders for each of the products within a folder named **Xtras** beside your projector:

- The "**Runtime**" folder for Impressario (remove the Resources folder to save space)
- The "**Runtime-vzd**" folder for INM VizionDB

Be sure that your own project files are in located in the proper locations:

- **PDF** files are placed within your movie's folder hierarchy, in the relative folder location that was used when you created your Searchable Library.
- The **INM VizionDB database file (.vzd)** is placed in the folder next to your movie
- The **Searchable Library Index file (.vxd)** is next to the INM VizionDB database file.

Make sure the path to the database is specified as a relative path in the INM VizionDB Database member properties dialog.

If you are creating a cross-platform CD, consult our Knowledgebase for articles on how to publish your projects for cross-platform CD delivery:

Test your project on each anticipated delivery platform before burning large quantities of CDs/DVDs.

Customization of the Searchable Library Solution

INM has provided a variety of pre-scripted tools and API's to accommodate typical searching needs. However, as project specifications vary widely, you may have specific needs that are not met directly by our built-in Searchable Library function, or other development tools.

If you have a specific project need that you are not able to accomplish using the tools provided, we may be able to develop a solution that is tailored to your specific needs. Contact INM Consulting Services if you would like more information on the type of customization we can do and for an estimate.

Custom document properties

If you would like to store additional information about each PDF document in the database, depending on the application you use to generate the PDF files, you may be able to create custom document properties. Refer to Adobe Acrobat's documentation and the documentation of the application that generates your PDF files for information on adding custom document properties.