Full Text Service
User Guide

Version 6.1

DOCUPHASE

Version 6.1   ---   1.23.2017

Table of Contents

Introduction

Welcome to Full Text

The Full Text product is a service that runs behind the scenes to extract and catalog significant words present in all of the document content stored in DocuPhase for designated Application cabinets.

With this extensive catalog of words, Full Text searching is different than searching for documents using index fields because full text search allows keyword searches across OCR-text of scanned images as well as Word documents, Excel spreadsheets, iForm documents, HTML files, text files and PDF files.

In DocuPhase a Full Text or Advanced searches may be combined with search criteria for DocuPhase index fields which allows the combined search criteria to pinpoint very specific information contained in any document.

Purpose and Assumptions

This manual is written for DocuPhase Administrator Users to prepare them to install, configure and use the *Full Text* product.

The following are assumptions made with regard to the reading audience for this guide:

- The reader has received at least one week of DocuPhase provided training.
- The reader has used the DocuPhase software product for a period of no less than one month.

Foundational Terminology

The following table defines terms which are needed to understand Full Text administrative processes.

Term	Definition
Full Text History	The Full Text History is a database table containing historical information regarding Full Text indexing activity for all documents.
Full Text Indexer	The Full Text Indexer is a command line application used by the Full Text Service.
Full Text Manager	The Full Text Manager is desktop application deployed to client site administrators. It is used to manage the Full Text Queue and Full Text Archive for individual applications
Full Text Queue	The Full Text Queue is a database table containing references to documents waiting for indexing.

Term	Definition
Full Text Service	The Full Text Service is a Windows service that monitors the Full Text Queue for documents that need to be Indexed.
OCR	Optical Character Recognition; the process of translating images of typewritten text into machine-editable text.
Trigger	Code that is automatically executed in response to certain events. Enabling Full Text adds a trigger to the associated application that monitors documents that are modified and places them back in the Full Text queue for re-indexing.

Client Support Services Contact Information

DocuPhase is committed to providing quality service and support for our customers. If you are experiencing difficulty with your DocuPhase software, please let us hear from you so we can help.

Client Support Services are provides as part of your Maintenance Program. Enhanced support programs are available upon request. The standard support feature set includes:

- Product Updates and Upgrades
- Telephone and Email support during local business hours
- Remote Connect Support during local business hours

You may:

- Contact us by email at support@DocuPhase.com.
- Reach us by phone at (727) 441-8228.
- Reach us by fax at (727) 444-4419.
- Find us online at www.DocuPhase.com/support.

Full Text Product Overview

DocuPhase Full Text Service provides the ability to search for specific words or strings of words across documents stored in a DocuPhase application. When performing a full text search, the requested words are highlighted and can be identified on a per document basis.

This advanced search capability for full text searches is possible because the Full-Text Service is able to initially scan all of the text in documents across many application virtual filing cabinets in DocuPhase and catalog these words to support efficient searches. The Full Text Service operation is performed automatically at scheduled times that are often at off-hours to not compete with normal production operations.

Full Text searching is different than searching for documents using index fields because full text search allows keyword searches across the text-content derived from scanned images, Word documents, Excel spreadsheets, HTML files, text files and PDF files. It can optionally include special content such as: annotations on image pages, item tags, special notes on DocuPhase items and even text in updated event history records for documents as well as in prior revisions of documents.

A full-text search may also be combined with an index search to pinpoint very specific information contained in any document.

Full Text Manager

The Full Text Manager component controls which applications are Full Text enabled and allows the system administrator to monitor the process of Full-Text OCR and indexing in real time.

Each DocuPhase cabinet can have Full Text enabled or disabled as well as include Document Revision copies or not when Full-Text word-indexing is enabled. Full Text is enabled at the application level by means of a checkbox within the Full Text Manager, shown below.

Once an application is Full Text enabled, the Full Text Service will query the database and determine whether it should perform a full-text scan and indexing pass on the documents within the specified application's repository.
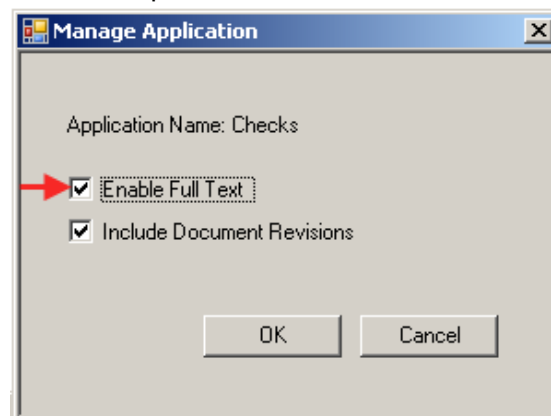
> **Note:** Full Text Manager is not responsible for performing document OCR; it is simply responsible for managing which applications have full text enabled.

Full Text Manager gives users the option to change the recognition engine type during processing. There are currently five different types of recognition engines with varying degrees of accuracy and speed.

> 🌀 **Note:** *Appendix C has descriptions of each recognition engine along with Full Text Manager which also allows the system administrator to view documents that are in the Full Text indexing queue and history database tables.*

The system administrator can browse directly to documents by double clicking on the selected row within the Queue or History tabs. The system administrator can also narrow their search for specific documents by Date and/or Application.

Documents move from the Queue tab into the History tab once Full Text Indexing has occurred.

Full Text Service and Full Text Indexer

Full Text Service is responsible for starting and stopping the Full Text Indexer application. Full Text Indexer uses the third party Scansoft API engine to OCR TIFF documents.

In order to properly process documents 24 hours a day, seven days a week, the Full Text Service starts and stops the Full Text Indexer after processing each page of a document. This method of deployment guarantees the stability of the Full Text Service.

The Full Text Indexer component performs optical character recognition on TIFF images and places the associated text within the FTSData table of Microsoft SQL Server.

Although Full Text Indexer only applies to TIFF images, Full Text searches within DocuPhase extend beyond this limitation as Microsoft SQL server natively supports a variety of document types for Full Text. The list of file types supported and their associated iFilters are described in Appendix B.

The Full Text capability can also be extended to PDF files through the installation of the Adobe iFilter on the SQL Server.

Enterprise Scalability

The Full Text Service is designed with an enterprise-wide view. Full Text Service employs record locking, thereby making it possible to run Full Text Service on multiple machines to effectively double or triple the processing capacity of full-text Indexing to scale its speed and processing capacity to meet the document volume demands of enterprises of any size.

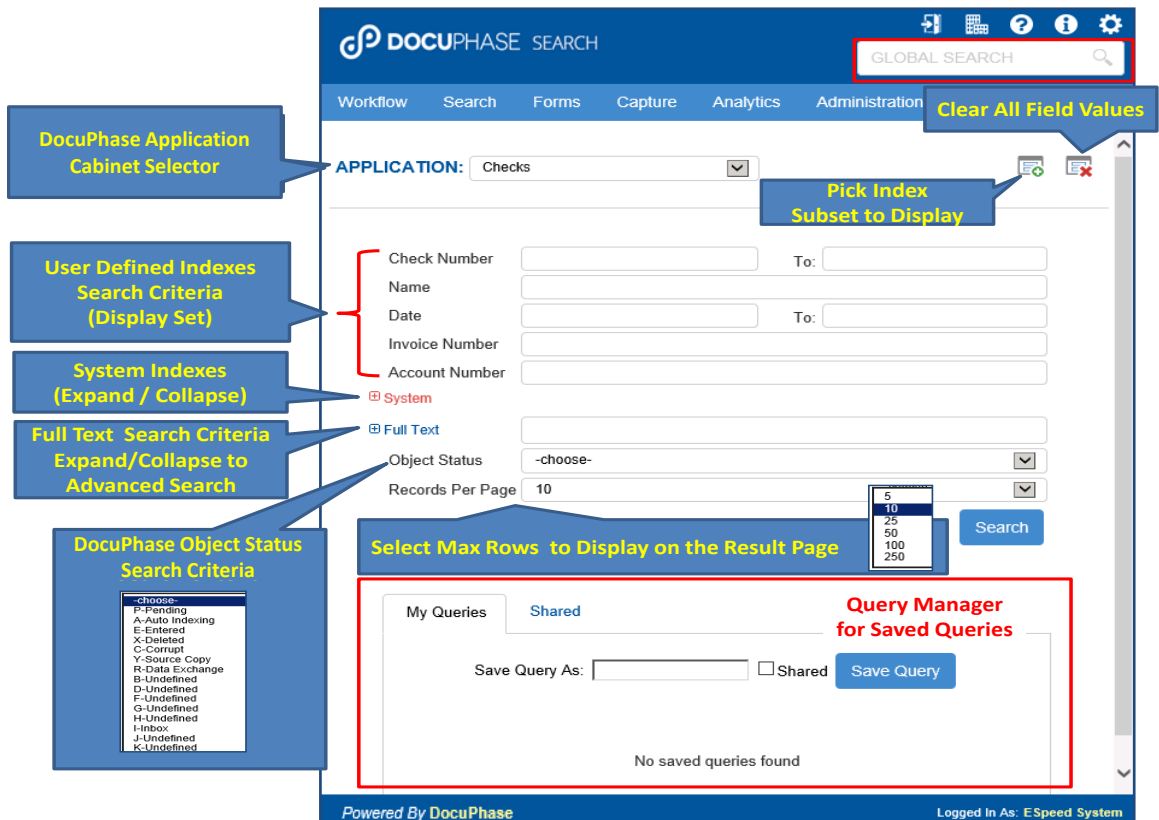> 🌀 **Note:** Each instance of the Full Text Service will require a license

Full Text Search Interface in DocuPhase

The Full Text product components that make the combined catalog of indexed words from documents across all full-text enabled Applications make it possible for DocuPhase users to perform advanced and precise searches for the documents they need within a specific application using their Full Text Search interface in DocuPhase.

Full Text search can be applied in DocuPhase in the following ways:

- Advanced Word-Search within an Application
- Advanced Word-Search Globally across all enabled Applications

The DocuPhase Search page shown below has annotations that show the features for both Global and Application document searches that include options to search using Full Text.

> **Note:** *Menus and fields displayed on the Search page may vary per user depending on permissions granted to each user by the administrator.*

Full Text Search

Full Text Search – Basic Advanced Search Mode: Advanced Search is available in two different modes.

The basic mode is labeled as 'Full Text' on the Search page and it will simply search across all of the following metadata fields for the selected Application cabinet:
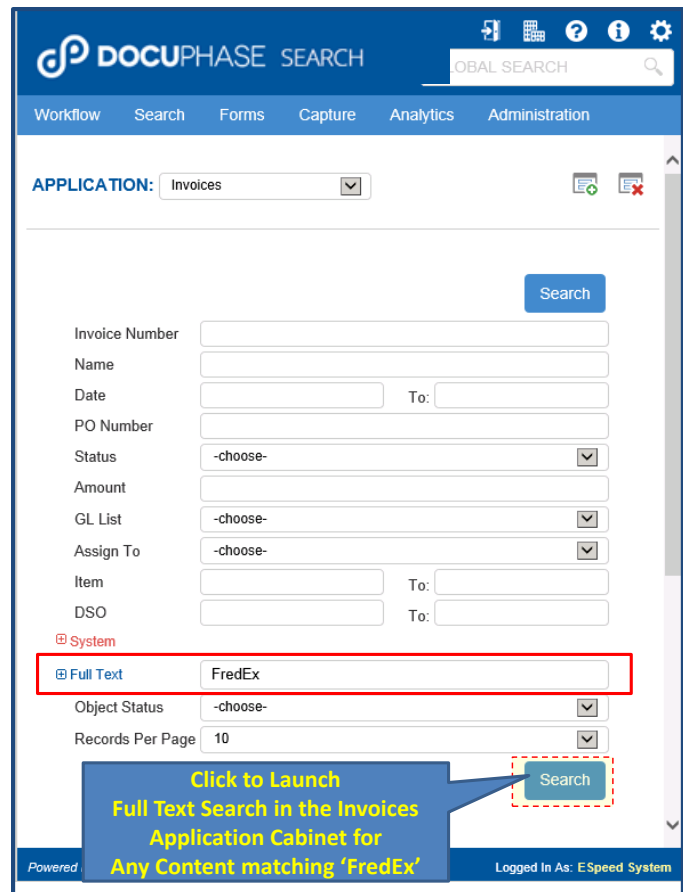
- **Document Text (also known as Full Text)** – search across full-text data that has been performed by the DocuPhase Full Text Service.

- **Annotations** – search across all annotation metadata stored with records. The ability to search for annotations is limited by the user's ability to view annotations.

- **Notes** – performs search of all notes stored within the system. The ability to search for notes is also limited by the user's ability to view notes.

- **Tag** – performs search of all tags stored within the system. The ability to search for notes is also limited by the user's ability to view tags.

- **Event History** – performs search of the event history of all records stored within an application.

The Full Text search argument can also be expanded (⊞)using the Expand/Collapse controls (i.e., ⊞ ⊟) to open the Advanced Search feature allowing separate full text search criteria to be applied concurrently to each category of metadata. See the next topic ***Advanced Search – Advanced Mode*** that follows.

The Full Text Search example, below, looks for the value 'FredEx' in any of the content categories for Advanced searches.

The advanced search option in basic mode gives the user the ability to search in an "OR" type fashion.

For instance, when a user enters the word "FredEx" into the Full Text Search field above, the DocuPhase search engine looks for the word "FredEx" in all of the Advanced option fields and displays all documents in the selected Application that meet the search criteria in any of the categories.

Note: Entry of a compound search criteria with multiple values such as Full Text = [FredEx Clearwater] would be still be evaluated as: "FredEx AND Clearwater" in each category where it is applied.

The "OR" type search would be evaluated across the available categories based on permission settings, as follows:

- Document Text contains: "FredEx" OR
- Annotations contains: "FredEx" OR
- Notes contains: "FredEx" OR
- Tag contains: "FredEx" OR
- Event History contains: "FredEx"

Note: Document Text is also referred to as Full Text and it, like the other service supported options, may or may not be installed or the Application or User permissions may not make it available for a Full Text or Advanced search.

> ⊙ **Note:**
> *The Full Text Mode and Advanced Search Modes are <u>NOT Case Sensitive</u> so it does not matter whether the text search argument(s) or target values are in capital or lower-case letters.*

If the DocuPhase repository contained one document with the word "FredEx" in a *Note* and another document had an *Annotation* with the word "FredEx", then the DocuPhase search engine would return two (2) records.

> *Note: Any document in the selected Application that is matched by multiple search criteria is only returned <u>once</u> as simply a matched document.*

> ⊙ **Remember:** *Unlike Global Search, Full Text Search and Advanced Search modes are <u>limited to searching only the current DocuPhase Application</u> selected on the Search Page.*

Compound Search Arguments

Six search clauses can be used in conjunction with Full Text and Advanced Search. These include the following topics:

- **AND Search,**
- **OR Search, Quoted Search,**
- **NOT Search,**
- **Wildcard Search, and**
- **Combination Search**

AND Search

By default, a User can enter a single word during a Full Text Search. The server then returns Documents that contain occurrences of that specific word.

However when the user enters two or more words as search criteria, the server returns only those Documents that include occurrences of all words contained in the search. The AND clause is implied between each word in a multi-word search phrase.

> *For example, performing a search with the search phrase 'White Orchid' will return only Documents contain both the words white and orchid in their content. A Document containing only the word 'white' or only the word 'orchid' would be excluded from the search results.*

OR Search

An 'OR' search is much different than an 'AND' search. The User must manually enter an OR (case sensitive) clause between each word in this type of search string for it to succeed. Without the 'OR' statements between words in the string, the string will be interpreted by default as having 'AND' clauses between consecutive words in the search string. When 'OR' is placed between search terms, Documents with occurrences of any word within the search string will be returned.

> *For example, searching for 'white OR orchid' returns Documents that contain either the word white or orchid or both.*

Quoted Search

A Quoted search allows the user to search for a specific phrase. This is different from an 'AND' search in that the order of the words in the search phrase are significant. A 'Quoted' search is performed by enclosing a worded phrase in double quotes. When a User searches for a quoted phrase, the server returns Documents that contain exact matches of the worded phrase.

> *For example, searching for '"the cat jumped over the fox"' returns only Documents that contain the string 'the cat jumped over the fox'. Documents containing all the words contained in the phrase, but not in the order quoted, will not be included in the search results.*

NOT Search

In a 'NOT' search, the User attaches a minus sign before a word to exclude any Documents that contain that word.

> *For example, searching for the term 'Reptiles -lizards' would yield all Documents that contain 'reptiles' but excludes Documents that contain 'lizards.'*

Compound Predicate Search

A combination search consists of a query with two or more search predicates.

> *For example, an AND, OR, and wildcard search can be combined in one query. For example, the query string: "white orchid OR flam* "is a perfectly acceptable query.*

Microsoft SQL Search ignores and disallows certain characters:

Punctuation marks such as colons (:) and semicolons are ignored.

- Single letters (a, b, c) and digits (0, 1, 2, 3) as well as symbols ($ % ^?) are also ignored.
- Commas are not allowed and you may not search for quotation marks (").
- Single quotes (') are recognized and treated as a character within a word.

Microsoft SQL Search ignores certain words. Words ignored by SQL search are listed in the Table below.

about	be	come	he	is	my	said	the	too	which
after	because	could	have	it	never	same	their	under	while
all	been	did	her	like	now	see	them	up	who
also	before	do	here	make	of	should	then	very	with
an	being	each	him	many	on	since	there	was	would
and	between	for	himself	me	only	some	these	way	you
another	both	from	his	might	or	still	they	we	your
any	but	get	how	more	other	such	this	well	
are	by	got	if	most	our	take	those	were	
as	came	has	in	much	out	than	through	what	
at	can	had	into	must	over	that	to	where	
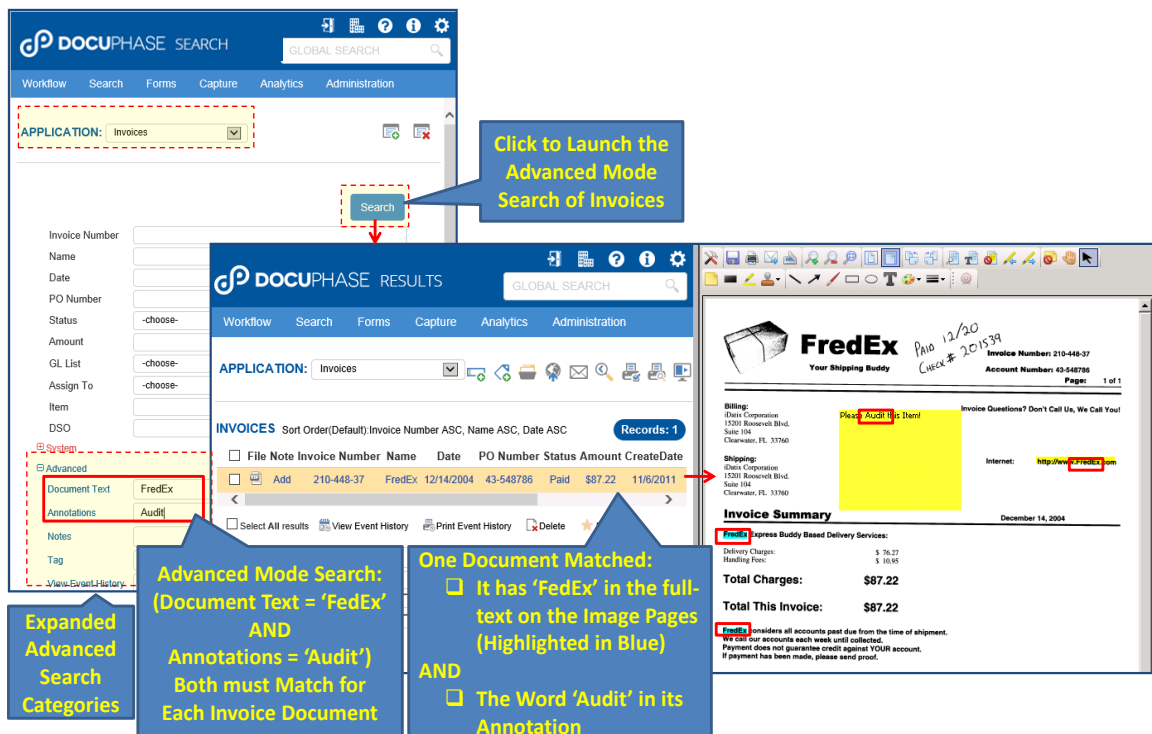
Advanced Search

Advanced Search – The Advanced Mode. The Advanced mode Search gives users the ability to search for specific individual values in each of the following categories by providing separate fields in which a similar or unique search criteria may be entered:

- Document Text
- Annotations
- Notes
- Tag
- Event History

The Advanced Search feature appears when the Full Text criteria is expanded by clicking its expand-control (i.e., ⊞) allowing individual search criteria to be entered for any or all of the displayed categories of metadata.

> *Note: The user may return to the advanced Full Text Search mode by clicking the collapse-control (i.e., ⊟) on the Advanced Search heading.*

> *Note: Entry of a compound search criteria with multiple values such as Full Text = [FredEx Clearwater] would be still be evaluated as: "FredEx AND Clearwater" in each category where it is applied.*

The advanced mode performs searches in an "AND" fashion similar to specific searches.

For instance, a user could enter the word "FredEx" into Document Text and the word "Audit" under Annotations to find all documents that have full text data "FredEx" in the document's image content and annotation(s) containing the text value "Audit".

This type of Advanced search is much more granular and specific than the Full Text advanced search mode since the category matches are AND-ed together.

The "AND" type search would be evaluated using the specified category search criteria fields based on permission settings, as follows:

- Document Text contains: "FredEx" AND
- Annotations contains: "Audit"
- Notes – not searched, since no criteria specified
- Tag – not searched, since no criteria specified
- Event History – not searched, since no criteria specified

  Note: Document Text is also referred to as Full Text and it, like the other options, may or not be installed or the Application/User permissions may not make it available for a specific Advanced Search.

> **Note:**
> *The Full Text Mode and Advanced Search Modes are not Case Sensitive so it does not matter whether the text search argument(s) or target values are in capital or lower-case letters.*

As illustrated in the previous example, only one document had both an Annotation with text containing the word "Audit" as well as one or more occurrences of the word "FredEx" in the document's text content obtained by OCR from the pages of this TIFF item available from the installed Full Text Service.

> **Remember:** *Unlike Global Search, Full Text Search and Advanced Search modes are limited to searching only the current DocuPhase Application selected on the Search Page.*

> **Important Note:** *Initially, NONE of the Advanced Global Search options (shown above) are set – Therefore, a Global Search will not return any matches until the appropriate options are set.*
>
> *Once set, these setting are retained for the user until they are explicitly reset (as shown above) and Save is clicked.*

The Full Text Manager Environment

This chapter offers an overview of the Full Text Manager environment. The Full Text Manager main screen consists of three areas:
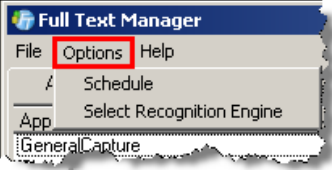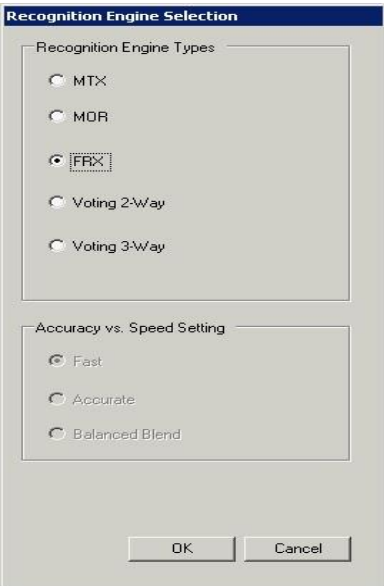
- Main Menu Bar
- Administration Tabs
- Status Bar
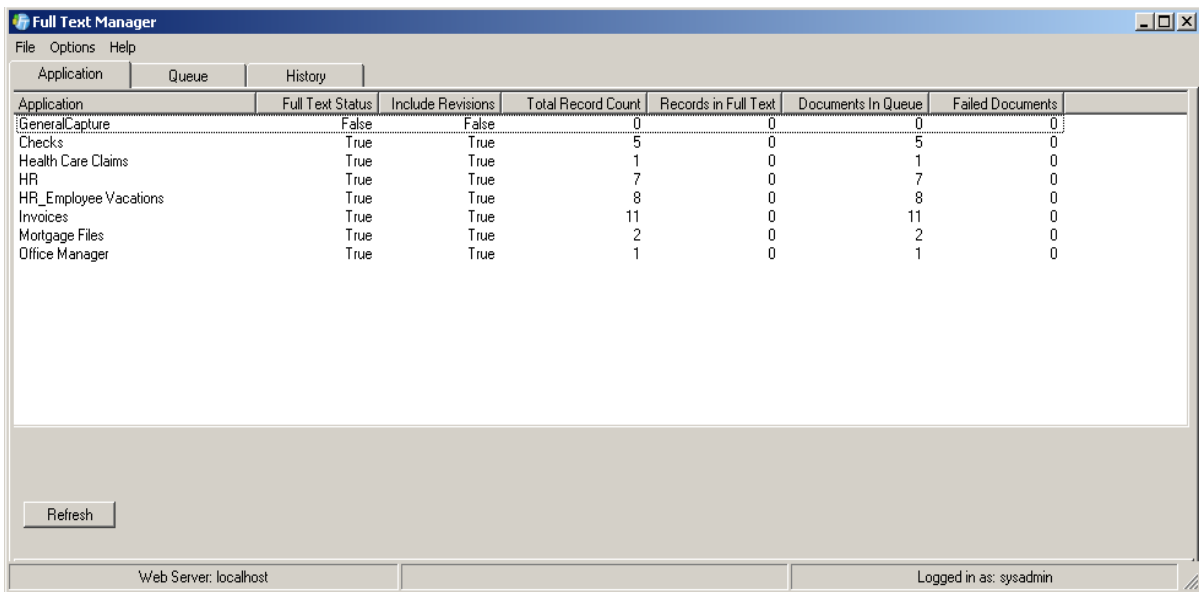
Main Menu Bar

The main menu bar contains three drop-down menus: File, Options, and Help.
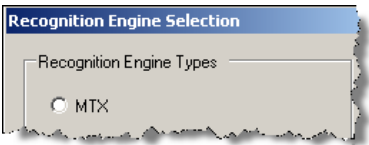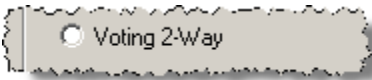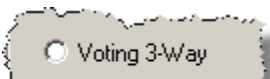
Element	Main Menu Bar Description
File Menu	The file menu provides the ability to logoff or exit the Full Text Manager application.
Options Menu	The Options menu consists of two options: Schedule and Select Recognition Engine. Use the Full Text Scheduling dialog to set start and end processing times. Use the Recognition Engine Selection dialog to select the recognition engine type and accuracy settings.

Element	Main Menu Bar Description
Help	The Help menu consists of one options: About. The About option provides a brief description of Full Text Manager by listing the product name, company name and version number

Recognition Engine Selection

The following element and description table has been included to enable the user to understand how to use each of the recognition engine types.

OCR Engine	Recognition Engine Selection Description
MTX 	• Fast selectable OCR engine. • Support for twelve (12) languages. • Supports a maximum of 64 zones on one image. Supports Omnifont, Draftdot9 and Draftdot24 filling methods. • Provides two (2) page-level accuracy and speed trade off settings including a combined Accurate and Balanced value and Fast. • Provides Checking Subsystem based correction.
MOR 	• Supports 119 languages. • Supports a maximum of 500 zones on one image. • Supports Omnifont, Draftdot24 and OCR-A filling methods. • Supports character training to achieve improved accuracy. • Provides three (3) page-level accuracy and speed trade off settings including Accurate, Balanced and Fast. • Provides Checking Subsystem based correction.

OCR Engine	Recognition Engine Selection Description
FRX ⊙ FRX	• Optimized for speed. • Support for 54 languages. • Supports a maximum of 2,500 zones on one image. • Supports Omnifont filling methods. • Supports character training to achieve improved accuracy.
Voting 2-Way ○ Voting 2-Way	• PLUS 2-way voting engine for accurate and fast machine print OCR capabilities.
Voting 3-Way ○ Voting 3-Way	• OmniPage 3-way voting engine for accurate and fast machine print OCR capabilities.
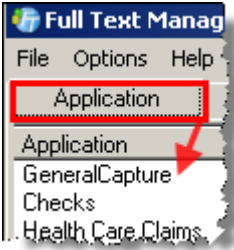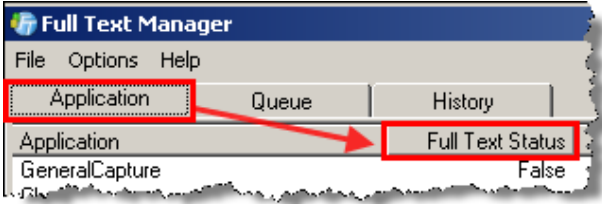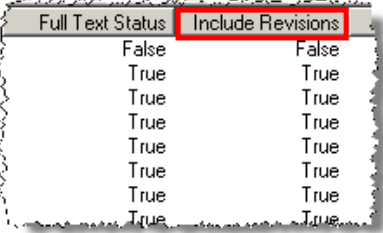
Administration Tabs
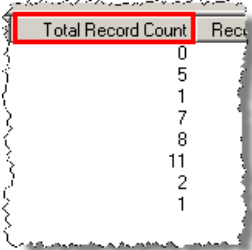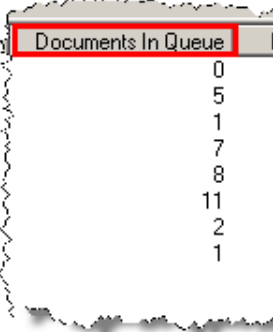
Application Tab
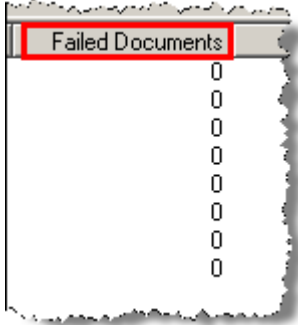
The Applications tab allows the administrator for Full Text to enable a specified group of documents, as defined by a DocuPhase application. Upon enabling an application for Full Text, the administrator is prompted for confirmation and required to select whether or not existing documents should be placed in the Full Text queue.
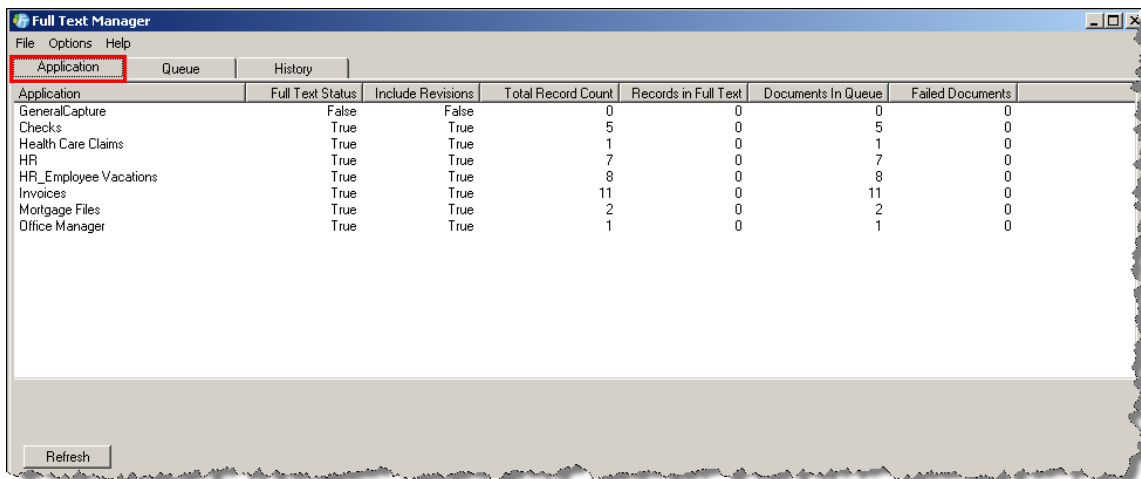
> **Note:** The system administrator is responsible for managing the Full Text Manager application. The Full Text Manager provides the administrator with three tab options to monitor document processing for Full Text Indexing.

Element	Full Text Manager Application Tab Description
Application Column	Provides a listing of all implemented applications in DocuPhase.
Full Text Status	Provides the ability to denote which applications you wish to enable with Full Text.
Include Revisions	Provides the ability to denote if Full Text Manager will also apply Full-Text processing to the document revisions.

Element	Full Text Manager Application Tab Description
Total Record Count Column	Presents the number of submitted items in an application that are either Full-Text Indexed or awaiting Indexing by Full Text.
Records in Full Text Column	Presents the number of records that have been Indexed by Full Text.
Documents in Queue Column	Presents the number of documents waiting to be Full Text Indexed.

Element	Full Text Manager Application Tab Description
Failed Documents Column	Presents the number of documents that failed during Full Text Indexing.
Refresh Button	Provides the ability to update Full Text Manager with the most current statistics for: Total Record Count, Records in Full Text, Documents in Queue, and Failed Documents.
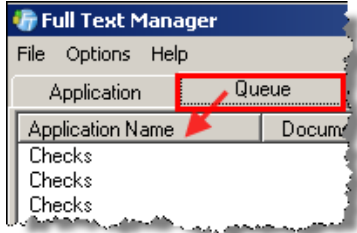
> *Note:* *For example, if Full Text is enabled for the "Checks" application, the user can elect to submit all existing documents into the Full Text Queue.*
>
> *All new documents are automatically submitted into the Full Text Queue. Processing time for existing documents may vary depending on the number of documents within the Application.*
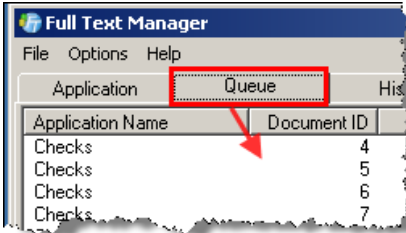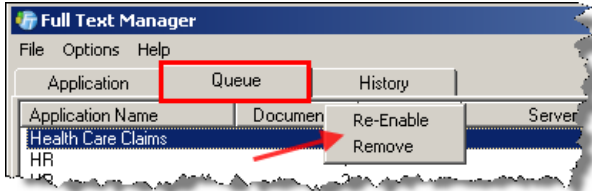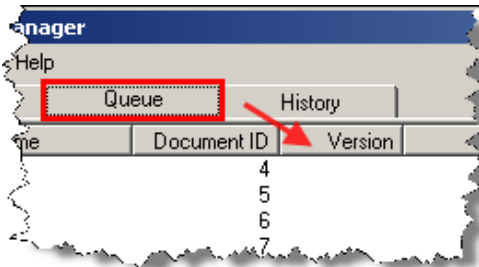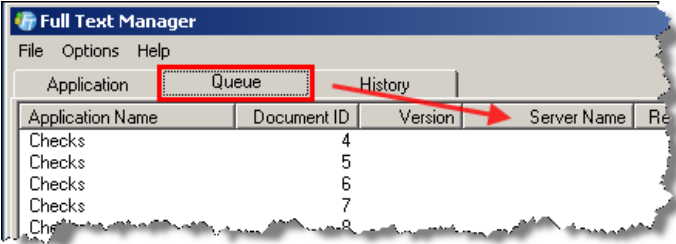
Queue Tab
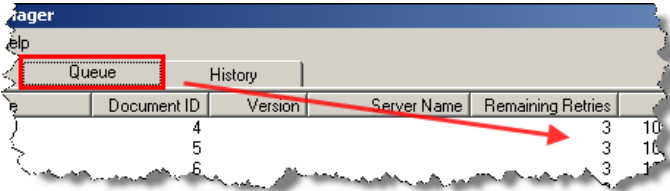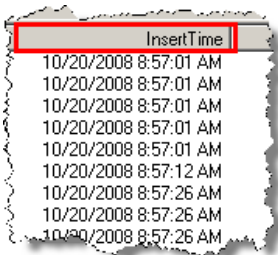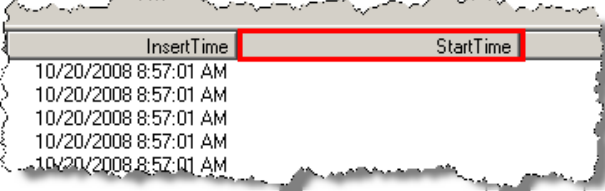
The Full Text Queue tab references the FTSQueue SQL table where document references are stored while waiting for indexing. The Full Text Queue allows the administrator to re-enable documents for indexing or delete them from the queue.
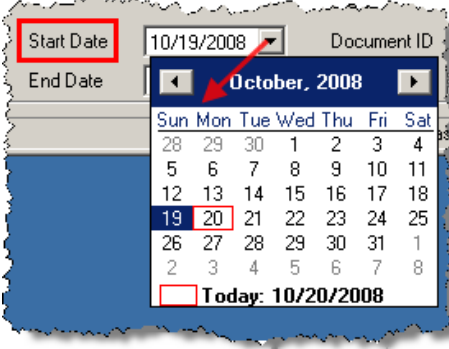
Documents are transferred from Full Text Queue to Full Text History when full text indexing is complete.
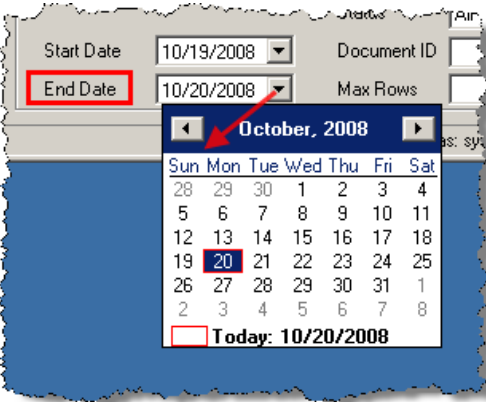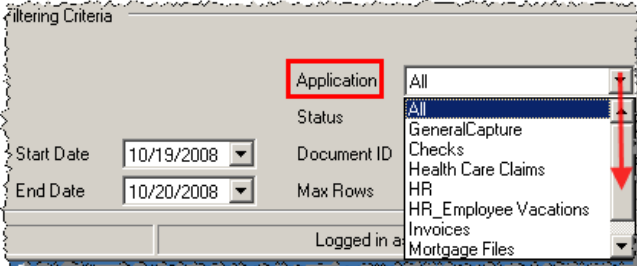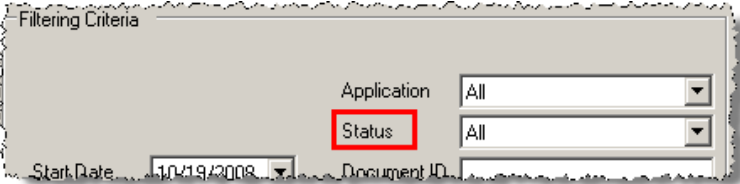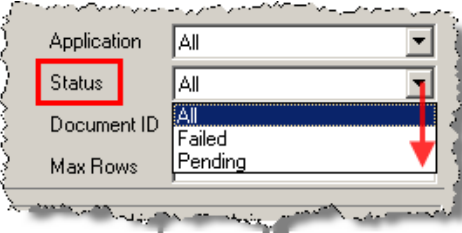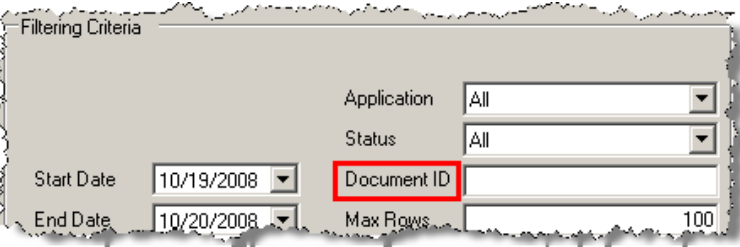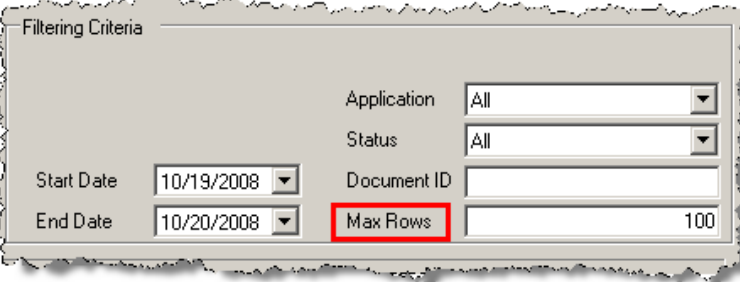
Element	Full Text Manager Queue Tab Description
Application Name Column	Displays an implemented and Full Text enabled application in DocuPhase.

Element	Full Text Manager Queue Tab Description
Document ID Column	Displays each document's assigned unique sequential number used by the system that becomes the primary key to identify the document. *Note: Additionally, you can right click and choose to re-enable or remove a document from the Queue.*
Version Name Column	Displays the full-texted document's version number. *Note: If the document version is equal to zero, then the version number will not be displayed.*
Server Name Column	Displays the name of the full-text server where the data is being accessed/processed.

Element	Full Text Manager Queue Tab Description
Remaining Retries Column	Displays a pre-determined number of tries to process the objects currently set to 3. *Note: By design, Full Text Service will only retry to perform full-text processing on a document three times.*
Insert Time Column	Displays the time the document was inserted into Full Text indexing.
Start Time Column	Displays the time the document started the Full Text indexing process.
Refresh Button	Provides the ability to updates the Queue Tab with the most current statistics.

Element	Full Text Manager Queue Tab Description
Filtering Criteria Start Date	Provides the ability to display a calendar to select a specific start date. *Note: The start date refers to the insertion time of documents.*

Element	Full Text Manager Queue Tab Description
Filtering Criteria End Date	Provides the ability to display a calendar select a specific end date. *Note: The end date also refers to the insertion time of documents.*
Filtering Criteria Application	By defaults to all Applications are selected. *Note: By selecting the drop down arrow, the administrator can specify a particular application.*
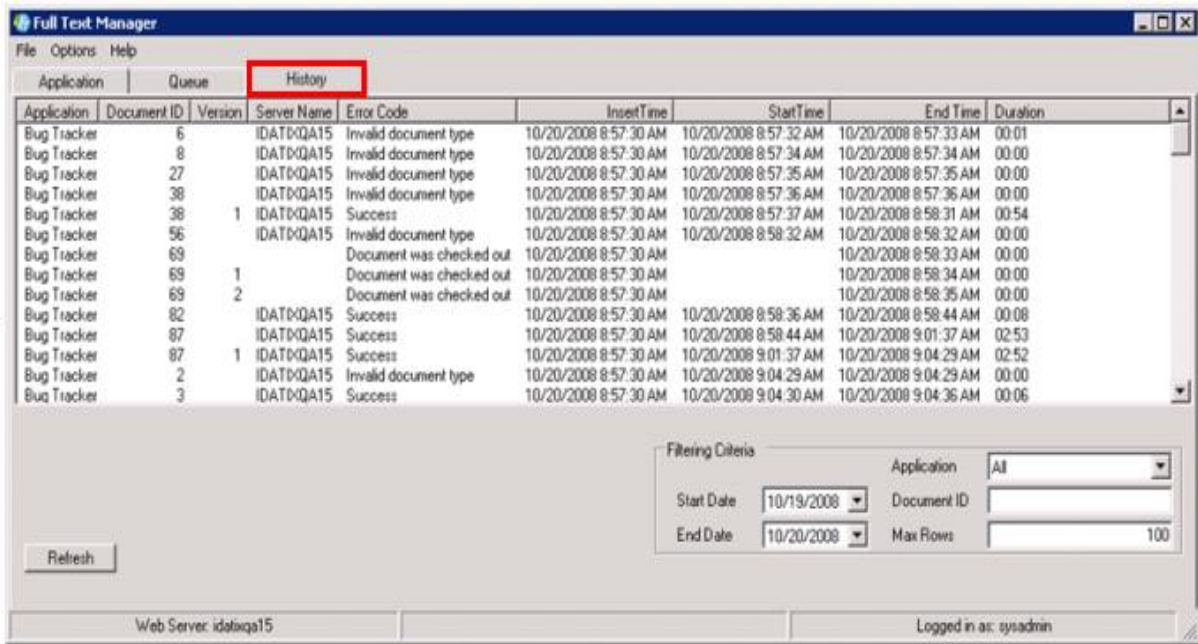
Element	Full Text Manager Queue Tab Description
Filtering Criteria Status	By default the filtering criteria is set to all. *Note: The Administrator may select the drop down menu and choose Failed, Pending or All status.*
Filtering Criteria Document ID	Provides the ability to define a specific document in the queue.
Filtering Criteria Max. Rows	By default the max rows are set to 100. *Note: The queue does not display the document being processed at the top of the queue.*

History Tab
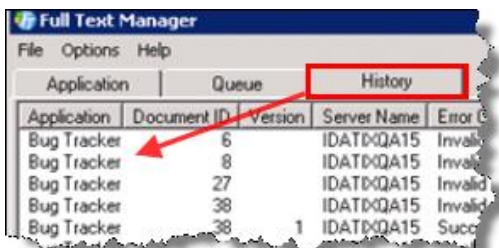
The Full Text History tab maintains a history of all documents processed by Full Text Indexer.
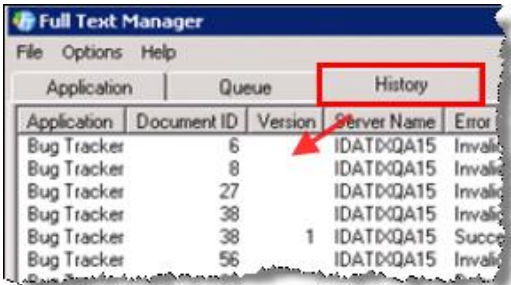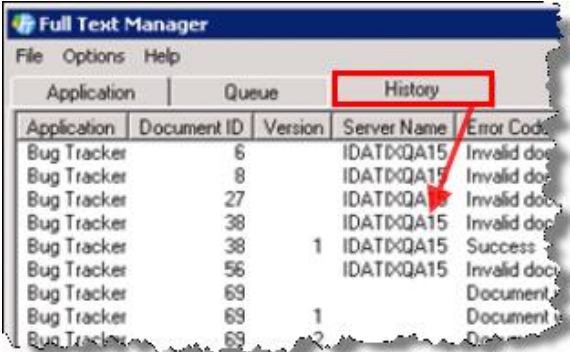
If a document in the Full Text Queue is successfully indexed, there will be one entry in Full Text History that shows success. Otherwise, *every* attempt at full text indexing a document is recorded in Full Text History.
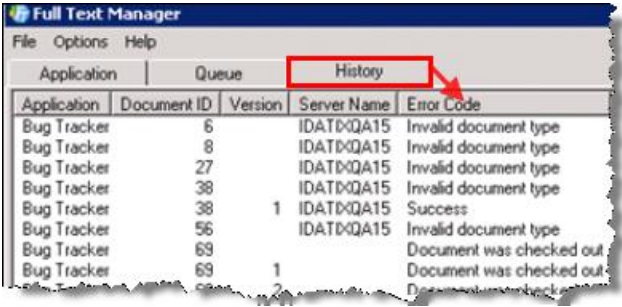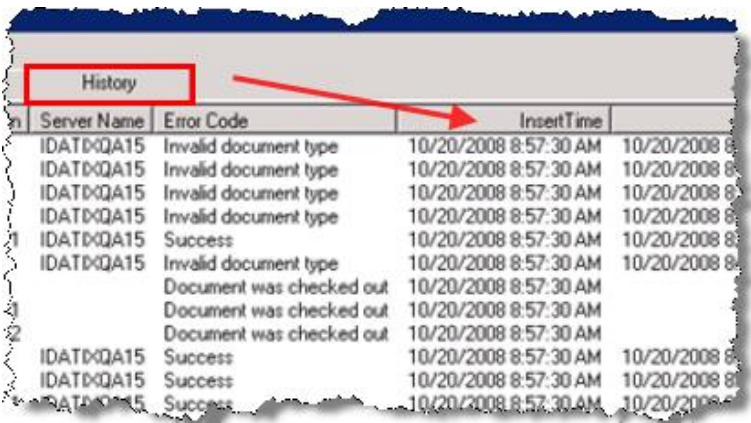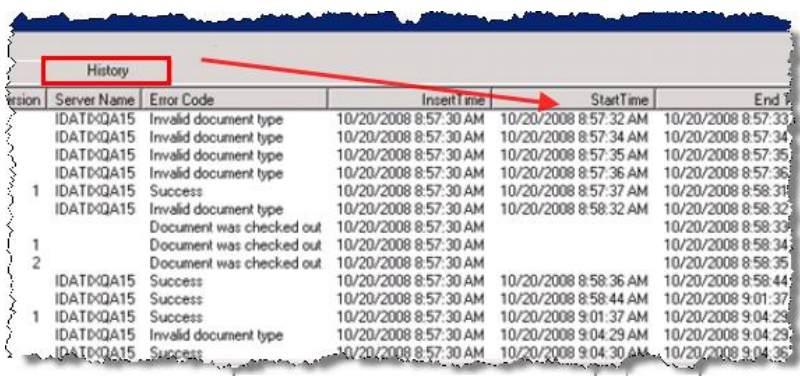
The maximum number of tries is set to three (3) per document. The Full Text Queue holds the document in an error state until it is manually deleted from the queue or resubmitted for processing.

Element	Full Text Manager History Tab Description
Application Name Column	Displays an implemented and Full Text enabled application in DocuPhase.

Element	Full Text Manager History Tab Description
Document ID Column	Displays each documents' assigned unique sequential number used by the system that becomes the primary key to identify the document.
Version Name Column	Displays the document version number. *Note: If the document version is equal to zero, then the version number will not be displayed.*
Server Name Column	Displays the name of the full text server where the data has been processed.

Element	Full Text Manager History Tab Description
Error Code Column	Displays the success or failure of a document.
Insert Time Column	Displays the time the document was inserted for Full Text indexing.
Start Time Column	Displays the time at which Full Text indexing began for the document.

Element	Full Text Manager History Tab Description
End Time Column	Displays the time at which Full Text indexing completed.
Duration Column	Displays the time Full Text Service took to process the document.
Filtering Criteria Start Date	Provides the ability to define a start date via a drop down menu calendar. *Note that the start date refers to the insertion time of documents.*

Element	Full Text Manager History Tab Description
Filtering Criteria End Date	Provides the ability to define an end data via a drop down menu calendar. *Note that the end date refers to the insertion time of documents.*
Filtering Criteria Application	Provides the ability to select the application you wish to see in the queue via the drop down menu.
Filtering Criteria Document ID	Provides the ability to select a specific document ID your wish to see in the queue.
Filtering Criteria Max. Rows	Provides the ability to select the maximum number of rows to be displayed.

Status Bar

The status bar, found at the very bottom of the Full Text Manager, displays the selected Database Server, Web Server, as well as the current user logged into Full Text Manager.

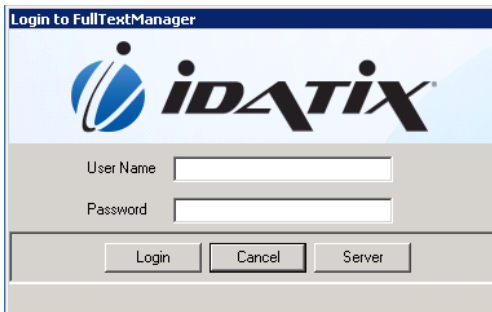| Web Server: localhost		Logged in as: sysadmin

Configuring and Scheduling Full Text

Use the following steps to log into the Full Text Manager and begin processing.

1. On your Desktop, select the **Full Text** Icon.

> *Note: The Login dialog is displayed.*

2. Enter a valid **User Name** and **Password**.

3. If this is the first use or a new Server needs to be defined, click the **Server** button (above)
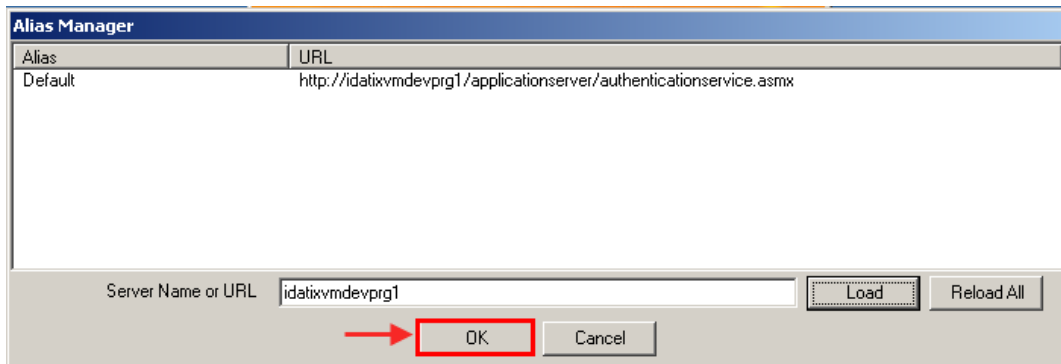
4. Then Select the **Manage** link.

> *Note: The Alias Manager Dialog displays.*

5. Define the **Server** information.

6. Select **Load**.
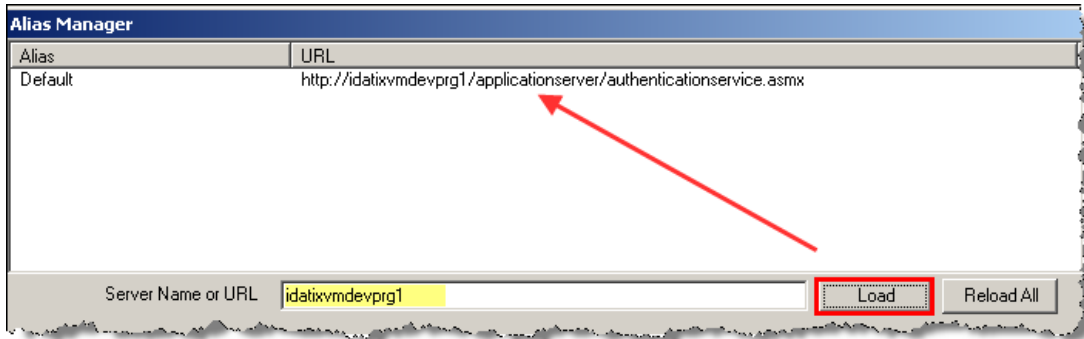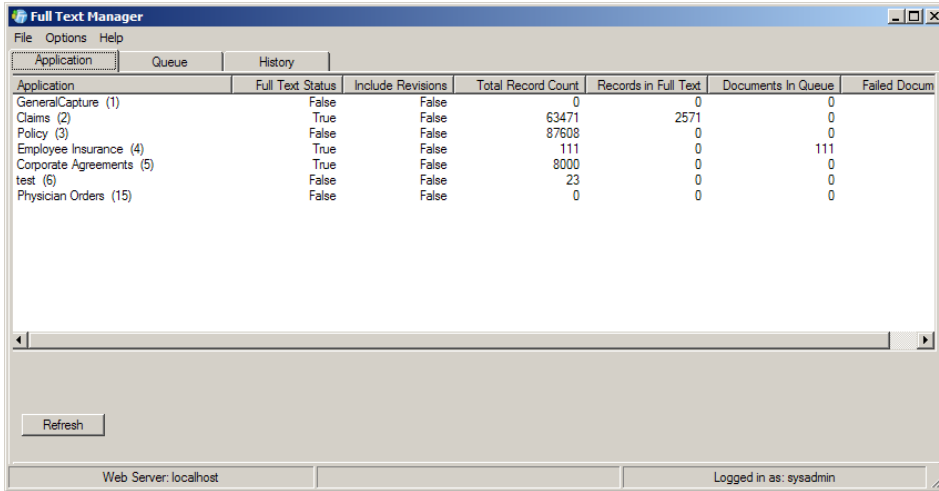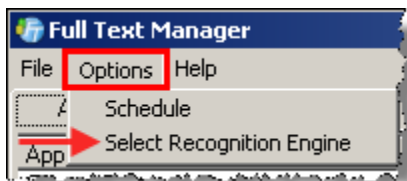

7. Select **OK**.

8. Select **Login**.

Note: The Full Text Manager application displays, shown below.

The Full Text Manager window displays with three administration tab options. A list of all implemented applications currently found in the system display.
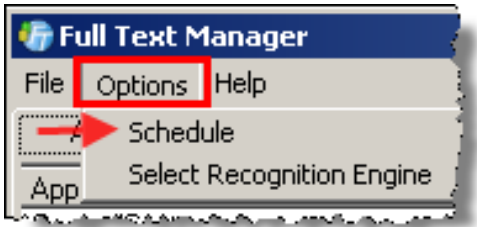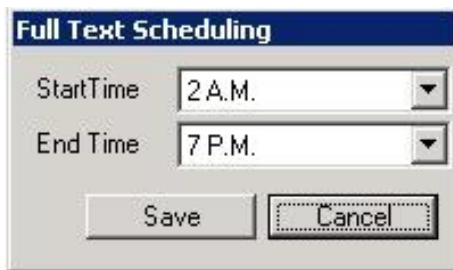

9. Set the Recognition Engine by selecting **Options > Recognition Engine Selection**.

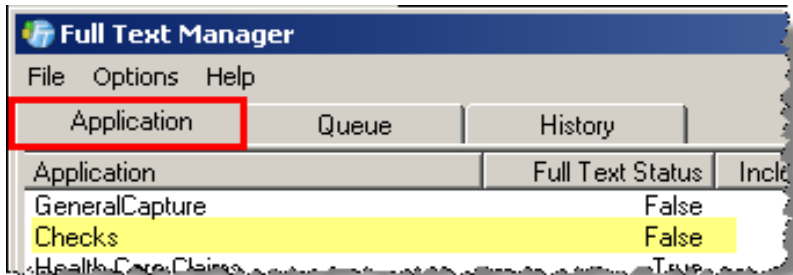10. Set the Full Text Schedule for processing times by selecting **Options>Schedule**.

11. Click **Save** to set the full text schedule.
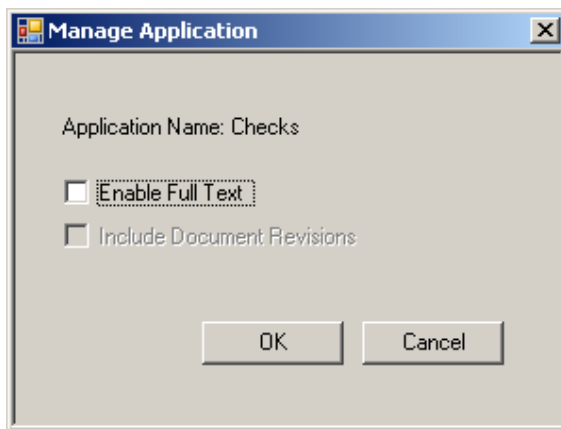
Enabling an Application for Full Text

To enable an application from the Full Text Indexing list from the Application tab,
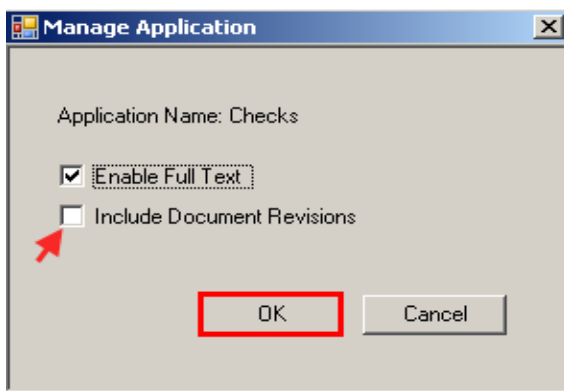
1. Click the Application.

Note: The Manage Application dialog displays.

2. Select the check box to Enable Full Text indexing for the selected application
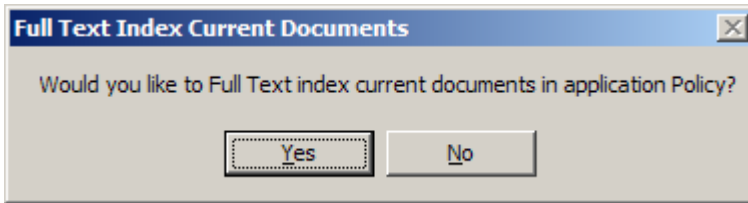

3. If you want to include older revisions of documents in the full text indexing, select **Include Document Revisions**.
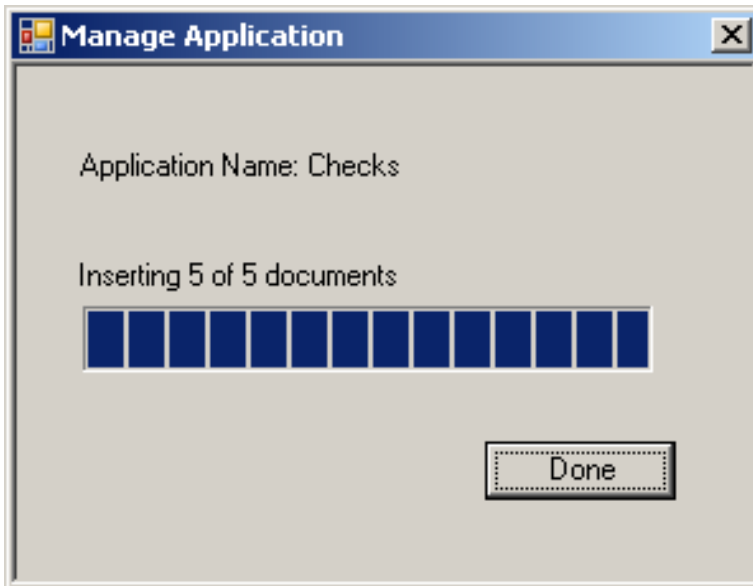
4. Click **OK**.

Note: The Full Text Index Current Documents dialog displays.

5. Select **Yes** to add the existing documents to the Full Text Queue.

Full Text Index Current Documents

Would you like to Full Text index current documents in application Policy?

Yes   No

6. Select **Done**.

Manage Application

Application Name: Checks

Inserting 5 of 5 documents
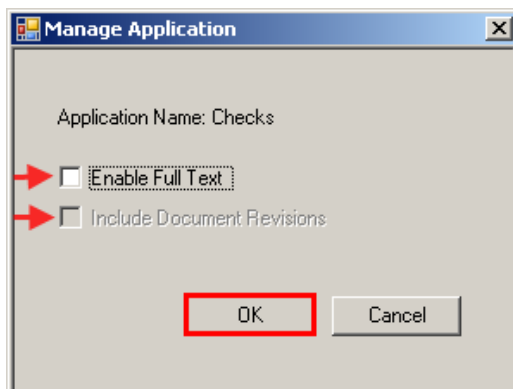
Done

Disabling an Application for Full Text

To disable an application from the Full Text Indexing list from the Application tab:

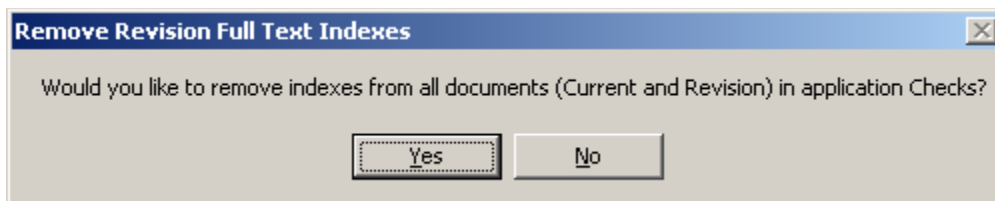**1.** Double click the Application you want to disable.
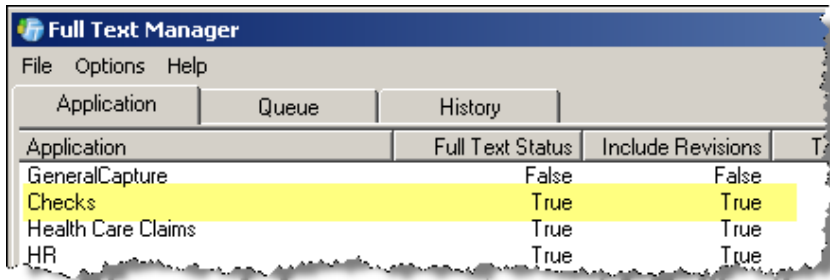
Note: The Manage Application dialog displays.

2. Uncheck the Include Document Revisions and the Enable Full Text check boxes.
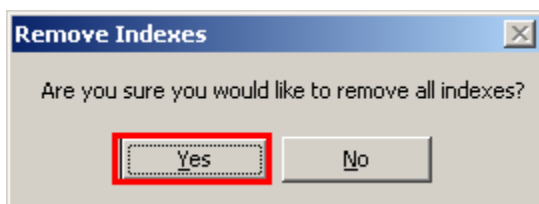
Select **OK**.

Note: The Remove Revision Full Text Indexes dialog displays.

3. Select **Yes** to remove the full text data from the database.
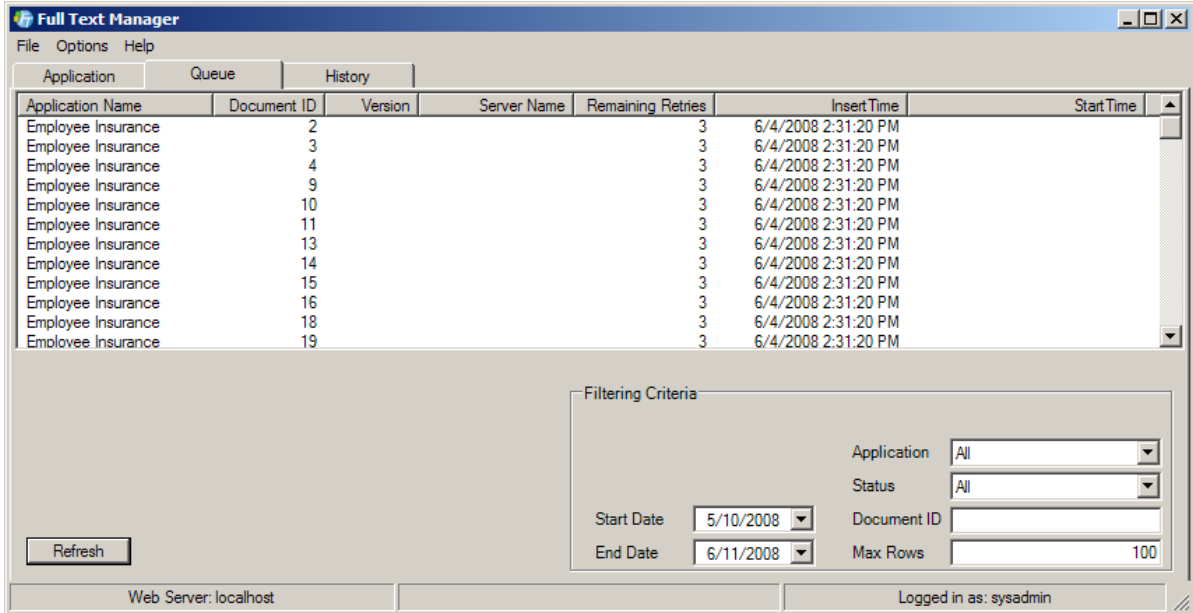

Note: The Remove Indexes dialog displays.

4. Select **Yes**.

Note: Full Text Service has now been removed from the selected Application.

Managing Applications in the Full Text Queue

1. Select the Queue tab to display the Queue window. Enter specific data in the Filtering Criteria area to narrow down the documents are pending in the queue.
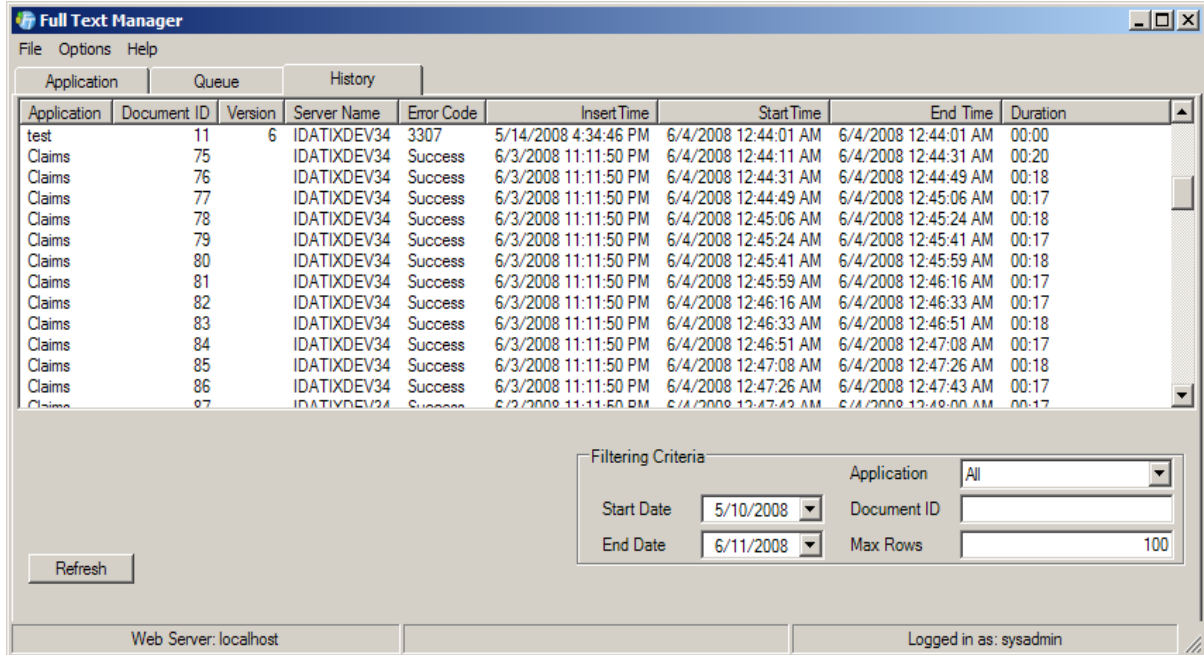

Note: The Application and Status field has a drop-down capability to display all applications entered in the system.

2. To set the Start and End Date, use the calendar to make your date selections.

3. Click the **Refresh** button to display the filtered results.

Viewing the Full Text History

1. Select the **History tab** option to display the processing history for all applications and their associated documents that have been Full Text indexed.

2. You may enter Filtering Criteria to narrow down your history result set. Click **Refresh** to view the new results.

3. Select and Click on any document to view the document within DocuPhase.

Appendix A: Full Text Installation

Prerequisites & Minimum Requirements

The following Full-Text installation consideration topics are documented in the following locations:

- **Minimum Requirements:** See *DocuPhase Prerequisites Guide*
- **Prerequisites:** See *DocuPhase Prerequisites Guide*
- **Full-Text Installation from .MSI:**
  1. **Full-Text Manager**
  2. **Full-Text Service**

  See: *DocuPhase Installation & Upgrade Guide*,
  Topic: *Installing the Optional Components*.

Once the Full-Text Manager is installed, its icon appears on the desktop of the unit where it is installed.
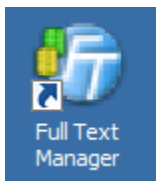
Full-Text Processor and Storage Considerations

The Full Text Service is extremely CPU intensive. It is not uncommon to see the Full Text Indexer application consuming close to 100 percent of the processor on a single CPU system.

In order to alleviate this problem, Full Text Manager has a scheduling option that allows processing during non-business hours. It is considered best practice to schedule Full Text Service to run on a nightly basis when CPU-cycle demand is generally at a minimum.

> **Comment:** CPU-*Processor consumption demand only occurs while the Full Text Indexer is performing OCR on a document.*

During the course of processing with Full Text Service, administrators will notice that the size of their database and/or repository will grow. There are two different methodologies for processing documents – TIFF and non-TIFF.

- The methodology for processing TIFF documents stores the OCR text data and coordinates within the database. This is different from previous versions of Full Text Service that created file revisions for each TIF document.

- The Full Text Indexer handles non-TIFF documents by simply placing a copy of the file within the FTSData table.

Appendix B - SQL 2005 Supported iFilters

Below is a list of iFilters which ship by default with SQL 2005. This list illustrates the file type index, and then the iFilter which will index it.

File Type	SQL iFilters
.obd	Office
.pot	Office
.doc	Office
.dot	Office
.log	Text
.rtf	Text
.ppt	Office
.txt	Text
.htm	HTML
.html	HTML
.xlb	Office
.xlc	Office
.xls	Office
.xlt	Office
.xml	XML

Appendix C – Recognition Engines

- **MOR OCR Engine**
    - Supports a maximum of 500 zones on one image
    - Supports character training to achieve improved accuracy
    - Provides three (3) page-level accuracy and speed trade off settings including
    - Accurate, Balanced and is Fast
    - Provides Checking Subsystem based correction

- **MTX (Mtext) OCR Engine**
    - Fast selectable OCR engine
    - Supports a maximum of 64 zones on one image
    - Provides two (2) page-level accuracy and speed trade off settings including a combined Accurate & Balanced value and is Fast.
    - Provides Checking Subsystem based correction

- **FireWorX OCR Engine**
    - Optimized for speed
    - Supports a maximum of 2,500 zones on one image
    - Supports character training to achieve improved accuracy

Appendix D – Inserting and Deleting Documents

This Matrix contains the Rules for inserting and deleting Documents from FTSQueue, FTSData and FTSCoordinate Data. Invalid Status conditions are handled accordingly with warnings and will always add documents to the FTSQueue when FTSCurrent is enabled.

Settings			Expected Results		
Revision Control	Revision Tracking	Full Text Current	Full Text Revision	Add New Docs to FTSQueue	Remove Revision FTSData as Documents are Checked in?
True	True	True	True	Yes	No
True	True	True	False	Yes	Yes
True	True	False	False	No	Yes
True	True	False	True	Invalid State	Invalid State
True	False	True	True	Invalid State	Invalid State
True	False	True	False	Yes	Yes
True	False	True	False	No	N/A
True	False	False	True	Invalid State	Invalid State
False	False	True	True	Invalid State	Invalid State
False	False	True	False	Yes	Yes
False	False	False	False	No	N/A
False	False	False	True	Invalid State	Invalid State