Inmagic PowerPack

Textbase Importer Service User's Guide

Contents

Contents

Chapter 1: Getting Started

The Importer is part of Inmagic® PowerPack™, a set of three Windows services: Checker, Updater, and Importer. You can use the PowerPack services on Inmagic® DB/TextWorks® textbases to help automate management and maintenance tasks.

Note: The Importer is not included with PowerPack Lite.

Audience for this Guide

This guide is supplied as part of the Inmagic PowerPack, which runs on the non-SQL platform. If you have **Inmagic DB/Text for SQL**, please see the *Inmagic DB/Text for SQL Importer Service User's Guide*.

Learning about the Importer

The Importer is a Windows service that processes imports in the background. The Importer looks for source files in one or more specified folders, loads the files into the specified textbase(s), then either deletes the source files or moves them to another folder.

To automate imports, and determine how they will occur, an administrator defines one or more import profiles per textbase. Each import can result in records being added, modified, or deleted. For example, one profile could import XML files in order to add or modify records, and another profile could import Inmagic Tagged Format files in order to delete records.

The Importer can load the following types of files:

- Fielded text files in Inmagic tagged, delimited ASCII, or XML format.

- The full text of documents, such as Microsoft Word files or other formats.

Import profiles for each target textbase determine how imports will occur. Each profile specifies:

- One source folder, to indicate the location of the import files.

- The format and parameters of the import files.

- Whether records will be added, modified, or deleted during the import.

- Whether source files should be moved or deleted after the import.

Note that the Importer service must be running in order for imports to occur.

Opening the Configuration Program

To configure the Importer, choose **Programs>Inmagic Applications>PowerPack>Importer** from the Windows Start menu. For instructions, see *Chapter 3: Configuring the Importer Service*.

Note: Whenever you make a configuration change, you must restart the service in order for your changes to take effect.

Starting and Stopping the Importer Service

The Inmagic Importer service must be running in order to perform imports.

To start or stop the Inmagic Importer service:

1. In the Windows Control Panel, open the Administrative Tools folder, and double-click **Services**. If you prefer, you can choose **Start>Run** and enter **services.msc**.

2. Right-click the Inmagic PowerPack Importer service.

3. Choose **Start** or **Stop**.

4. By default, the service is set to run **Manually**. If you want the service to start automatically whenever the system is rebooted, set the service Startup Type to **Automatic**.

5. Close the Services and Control Panel windows.

If you stop the Importer while it is processing a file, it keeps track of where it left off, and will resume processing the next time you start the Importer.

Be sure to stop the Importer service before rebooting the machine where the service is installed.

Chapter 2:   Preparing Files and Folders

This chapter describes the supported file formats and explains where to place files for import. It also describes additional folders that need to be set up before you start importing files.

Setting Up Import Folders

Before using the Importer, you need to create folders for the following purposes:

- **Source files**: Folders that contain files for import. A separate folder is required for each combination of textbase + file type + operation.

- **Exception files**: Location where the Importer should create files containing rejected records.

- **Processed source files**: Location where source files will be moved after import (if you choose to move, rather than delete, source files).

The account being used by the Importer requires full access to all of these folders.

Source File Folders

Before using the importer, you must create at least one (and usually multiple) source file folders. Files to be imported can either be manually copied to the specified folders or a knowledge agent can be set up to copy files there automatically.

Separate folders are required for each combination of textbase + file type + operation.

Use the following specifications to determine which folders to create:

- **Important!** Never specify the permanent directories where source files reside, because the Importer will either delete or move the files after the import. To retain your original documents, copy them to a different folder, which will function as the source file folder.

- The account being used by the Importer needs full rights to any import folders.

- Each folder must contain source files of only **one type** (Inmagic tagged, delimited ASCII, XML, or document). A folder that contains *document* files may include different document formats.

- Each folder must contain source files for only **one textbase**. This is how the Importer knows which source files go into which textbase.

- Each folder must contain source files that will use a **common set of import parameters**. Addition and modification can occur together, but files that will result in the deletion of records must be in a separate folder.

For example, you might set up the following import folders:

Location	Contents
C:\SALES\TAG\ UPDATE	Inmagic tagged files that will add or update records in the SALES textbase.
C:\SALES\TAG\DEL	Inmagic tagged files that will delete records in the SALES textbase.
C:\LIBRARY\XML\UPDATE	XML files that will add or update records in the LIBRARY textbase.
C:\LIBRARY\ DOC\UPDATE	Microsoft Word and Excel files that will be added as new records or replace existing documents in the LIBRARY textbase.

Exception File Folders

Before using the Importer, create an exception file folder for each import profile. An exception file is a text file that contains records rejected during an import. An exception file will be created only if records are rejected during the import. You can edit the exception file to make corrections to the records, and then reimport the corrected file.

- You can use any folder **except** a source file folder.

- Best practice is to create an exception file folder for each import profile, so you know where to move the files to be resubmitted after you fix the problem, and to use descriptive folder names. For example, if you have two profiles for the SALES textbase, one for new records and one for updating existing records, you could use exception folder names such as SalesExceptionsNew and SalesExceptionsUpdate. Or you could create subfolders, such as C:\SALES\ExceptionsAdd and C:\SALES\ExceptionsUpdate.

- The account being used by the Importer needs full rights to the specified folder(s).

For more information, see *Using Exception Files*.

Processed Source Files

When you create an import profile, you will indicate how you want sources files to be handled after processing. The Importer can either delete the source files or move them to another location. (The move or deletion occurs regardless of whether the files were successfully imported.)

If you want the files to be moved, you should create one or more destination folder(s) for the processed files. If a file with the same name already exists in a specified destination folder, the source file will be deleted rather than moved. Best practice is to periodically clean out the specified destination folder (back up the files if needed, then delete them).

Note: The account being used by the Importer needs full rights to any specified destination folder(s).

Accessing files and textbases on other machines

Ideally, the Importer should be installed on the file server containing the drive where the textbase files reside. This reduces network traffic because information does not need to be moved across the network.

However, if any of the following is true:

- The textbase that you plan to import into is located on another machine, or

- One or more of the folders specified in the Import Profile Wizard is located on another machine, or

- You are using a thesaurus as a validation list and the thesaurus textbase is located on another machine. **Note:** If any textbase uses a thesaurus as a validation list, the thesaurus textbase should either be in the same folder as the textbase, or the connection should be specified using the UNC filename, and the account used by the Importer must be able to see the thesaurus textbase files.

Then you must do both of the following:

- Set the Log On Properties for the Importer service to use an account that has full access to all the files or folders that reside on other machines, including the textbases in the [Textbases to Service list](). For example, use the same account as the one used to add the textbases to the list. Note that a thesaurus textbase only requires read-access, unless you use the .SLT file or enhanced logging on the thesaurus textbase.

- Use the UNC file name (for example, \\SERVER\VOL1\LIBRARY\CATALOG) instead of a drive letter (G:\LIBRARY\CATALOG) when specifying a textbase or folder name.

Obtaining Source Files

To obtain or generate files for import, use any method that produces a file in one of the supported formats. The most common methods are described below.

Download bibliographic data in XML or MARC format.

Use the originating application to export content to one of the supported source file formats. For example, use the DB/TextWorks Dump Textbase or Export feature to create a file in Inmagic tagged format. Or use a spreadsheet application to export content to a text file with field names in the first row.

You can also create a text file by hand, although this can be time consuming.

Application developers can use development tools such as Perl scripts or Visual Basic to create knowledge agents, which are programs or utilities that obtain information from a particular source. The knowledge agent should write the information to one of the supported formats, then put the file in a location where the Importer can find it. For example, a database administrator could write a program that exports data from a proprietary database to the Inmagic tagged format and places the file in the designated source file folder.

Tip! If the target textbase has an Automatic Date field (for creation or modification date), you can easily find records that have recently been added or changed. This is especially useful if you have a knowledge agent collecting information, and you would like to find out "what's new" periodically.

Source File Formats

The Importer service can import documents or fielded text files. Use the Import Profile Wizard to indicate the file type and specify how the files should be loaded.

Document Files

- A document file is any binary file, such as a Microsoft Word document or Excel spreadsheet, or any non-fielded or non-tagged text file.

- The entire text content of the incoming file will be placed in one field in one record. Note that an IFilter capable of extracting the text in the document is required on the Importer computer for this to work. Please see *IFilters* in the Glossary.

- Lines can be any length.

- You can include document files of differing formats in the same folder for import.

Fielded Text Files

- Fielded text files must be in one of the formats described below: delimited ASCII, XML , or Inmagic tagged.

- Use a different source folder for each file type. Do not put multiple file types in a single source folder. For example, put all delimited ASCII files in one folder and Inmagic tagged format files in a separate folder.

- One text file may contain one or many records.

- The DB/Text product line, including the Importer, accepts text using a single-byte character encoding such as ISO-8859-1 or code page 1252. This supports English and Western European languages, for example.

Delimited ASCII Format (fielded text)

- Files must meet the requirements stated above for Fielded Text Files.

- Use any standard characters to specify record separators, field separators, entry separators, and comments.

- Fields in the import file may be in any order, or you can put field names in the first row.

- Lines can be any length.

- For more information, see *Delimited ASCII Format*.

XML Format (fielded text)

- Files must meet the requirements stated above for Fielded Text Files.

- A transform file (XSL) is required to convert the incoming XML to a format suitable for import. You can either use the default transform, which is generated, or supply your own.

- Lines can be any length.

- For more information, see *XML*.

Inmagic Tagged Format (fielded text)

Inmagic tagged files are fielded text files in a format specific to Inmagic textbases. To create an Inmagic tagged file, use the DB/TextWorks Dump Textbase or Export feature.

Example

```
! This is a comment line (ignored during an import).
! This is the first record
AUTHOR Smith, Jane
SUBJECT Example of Inmagic tagged format
; Use a semicolon for each additional field entry.
; This is the third entry in the SUBJECT field. Continuation lines start
 with a space or a tab, as shown here.
> A line starting with a "greater than" sign indicates
a paragraph break to be retained within the data. A dollar sign on a line by
itself indicates end-of-record.
$
! This is the second record
AUTHOR Twain, Marcus
TITLE Huck Sawyer
"PUBLISHER NAME" Use quotation marks for field names that include spaces.
$
```

Characters at the beginning of a line indicate the items described in the following table.

Character	Description
!	Comment line (ignored during an import).
Field name	Field names can start with and include letters, digits, hyphens, and underbars. Field names that include spaces must be enclosed in quotation marks ('single' or "double"). Leave a space after the field name, then start the field content. Capitalization in field names is ignored.
;	Use a semicolon (;) at the beginning of a new line to indicate additional entries.
>	Line break to be retained in the data for previously named field.
Space or Tab	Continuation line for previously named field (permits wrapping).
$	A dollar sign on a line by itself indicates the end of a record.
other	Any other character at the beginning of a line results in an error, which causes the record to be rejected.

If you choose to create the file by hand, using a text editor, follow these additional guidelines:

- A file can include any number of records. Each record includes at least one field name followed by one or more entries. Do not include a field more than once in a record.

- Do not include a space or tab before a multiple entry, paragraph break, comment line, or end of record character (; > ! $). If you do, the character will be interpreted as text.

- Only include the fields that you want to affect. Fields can appear in any order (they do not have to match the textbase order.)

- A line cannot exceed 65,534 characters. To force a line wrap, press **Enter** and begin the next line with a space or tab. This is called a continuation line.

- Blank lines and comment lines are ignored during an import.

Chapter 3: Configuring the Importer

To configure the Importer, use the Importer Settings dialog to specify import options and define at least one import profile for each target textbase.

Inmagic Importer Settings

To configure the Importer

1. Choose **Start>Programs>Inmagic Applications>PowerPack>Importer**.
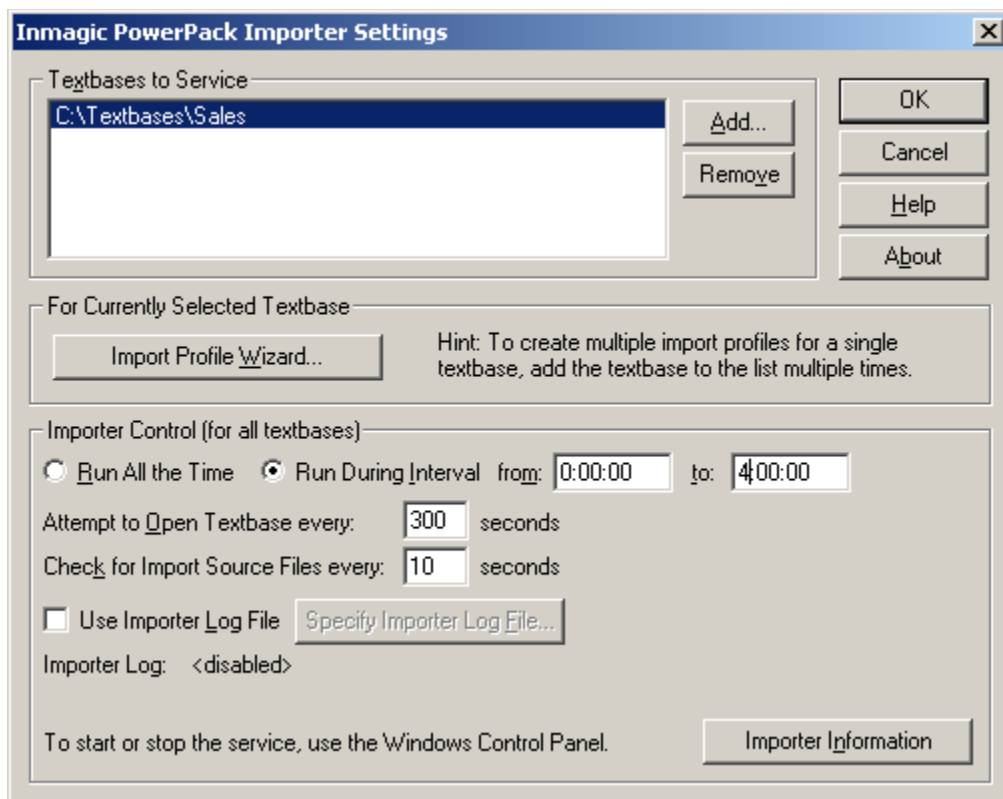
2. Click the **Add** button and select a textbase to add to the Textbases to Service list. If you add a textbase that has passwords, you must supply the Master password when prompted or you will not be able to use the textbase.

   Note: If the textbase is on a different machine than the Importer, use the UNC filename instead of a drive letter (\\SERVER\VOL1\LIBRARY\CATALOG). For more information, see *Accessing files and textbases on other machines*.

3. If you plan to create different import profiles for the same textbase, add the same textbase more than once. The Importer will process all textbases in the list simultaneously, so order is not important.

4. Select a textbase in the **Textbases to Service** list, click the **Import Profile Wizard** button, and create an import profile for that textbase. Repeat for each textbase (or textbase instance) in the list.

   Note: For instructions, see *Creating an Import Profile*.

5. Specify the **Importer Control** settings, which affect all textbases in the list:

- **Run All the Time.** Select this option to start running the Importer when you start the service in Windows (or immediately, if the service is set to Automatic). It will continue running until you stop the service in Windows, and will continually check the source file folders. Any services requiring exclusive access to a textbase (for example, the Inmagic Checker and most backup software) will not be able to open the textbase.

- **Run During Interval.** This is the recommended option. Select this option to run the Importer during the specified interval, so files will be imported only during the times specified. Specify a time that will not interfere with software that requires exclusive access to a textbase (such as the Checker and most backup software). If you are using Express Import, you should specify an off-peak time. If the Importer is in the middle of an import when the interval is over, it keeps track of where it left off. The next time the Importer is started, it will finish importing the file. Note that XML files are an exception-- the Importer will finish processing the current XML file, then stop. You must start the service in Windows to turn on the Importer (or set it to Automatic).

  Specify times using the HH:MM:SS format, using 24-hour time. For example, 0:00 starts at midnight (the default). 1:15 starts at 1:15 a.m. 12:00 starts at noon.

  Note: While the Importer is running, it holds open any textbases it is servicing. This can interfere with automated backups. To avoid conflicts, schedule imports for a different time interval than automated backups.

- **Attempt to Open Textbase every [nn] seconds.** Specify how often the Importer attempts to open a textbase to import records. If someone on the network has the textbase open using Immediate indexing, the Importer will not be able to open that textbase, but will try again after the specified interval. The default is 300 seconds (five minutes). Once a textbase is open, it remains open as long as the Importer is running (determined by the **Run All the Time** or **Run During Interval** setting).

- **Check for Import Source Files every [nn] seconds.** Specify how often the Importer attempts to load files. If you specify 10 seconds (the default), the Importer checks every 10 seconds to see if there are any source files in the Importer folders. If there are, it loads a file, then remains idle for the specified number of seconds.

- **Use Importer Log File.** Check this box to record activity in a log file (recommended). By default, the file is called IMPORTER.LOG. To change the name or location, click the **Specify Importer Log File** button. It is recommended that you always use a log file. For more information, see *Importer Log File*.

- **Importer Information.** Click this button to display and/or print information about the Importer, including the parameters defined for each import profile.

6. Click **OK**. This closes the dialog box and writes the settings to the services INI file.

7. If the Service is set to Automatic (in Windows), the Importer will run at the specified interval. If the Service is set to Manual, it won't run until you start it in Windows Services (see *Starting and Stopping the Importer Service*). **Important!** You must restart the Service in order for your changes to take effect.

Creating an Import Profile

An import profile defines how records will be imported into the selected textbase. You can create more than one profile for each textbase. (To create multiple profiles for the same textbase, use the Inmagic Importer Settings dialog box to add a textbase to the list more than once. Then create a different profile for each instance of the textbase.)

To create an import profile

1. In the Inmagic Importer Settings dialog box, select a textbase from the **Textbases to Service** list, then click the **Import Profile Wizard** button.

   The Import Profile Wizard appears.
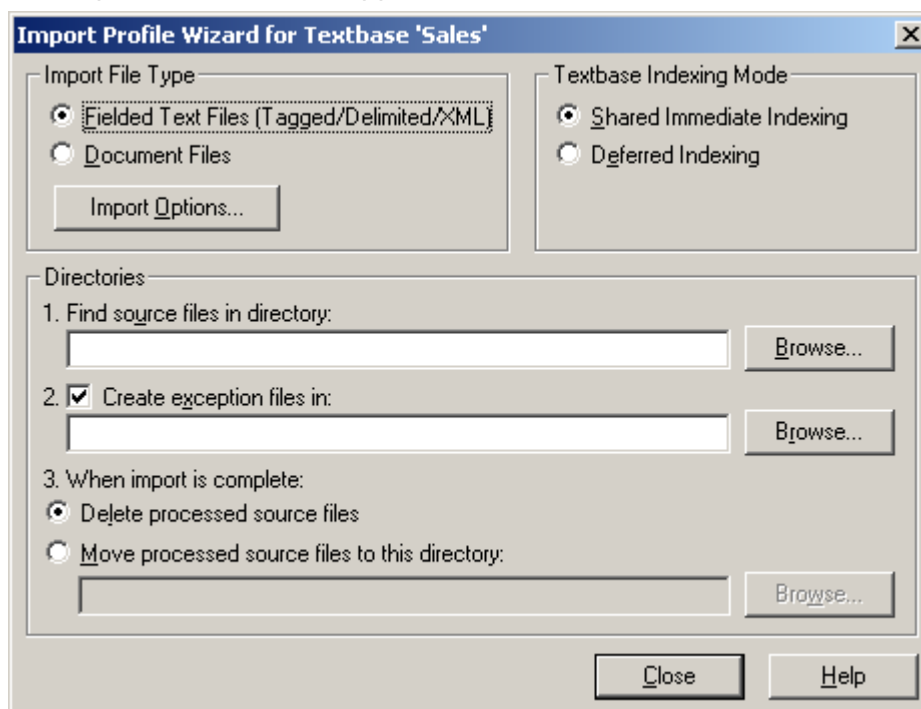
2. Select a **Textbase Indexing Mode** to determine when the Word and Term indexes will be updated.

   - **Shared Immediate Indexing.** Records are indexed immediately. During the brief period that indexing occurs (usually several seconds), other users of the textbase are temporarily denied access to operations that involve the indexes, such as searching or saving records. This is the recommended setting if the Importer is set to run at an off-peak time.

   - **Deferred Indexing.** Hold changes in a queue until deferred updates are posted, either by the Updater or by using DB/TextWorks. This allows many people to share the textbase on a network without indexing delays. However, until deferred updates are posted:

     - New unposted records will not appear in a search result (although a command query "getdeferred new" will retrieve them).

     - Unposted changes to existing records will not appear when records are displayed, printed, or sorted, unless you use **Tools>Options>General** in DB/TextWorks to specify otherwise.

3. **Find source files in directory**: Specify where the import files are located. If the files are on another machine, use the UNC filename (\\SERVER\FOLDER) instead of a drive letter. The Importer needs full rights to the specified folder.

   Important! Never specify the permanent directories where source files reside, because the Importer will either delete or move the files after the import. To retain your original files, copy them to a different folder, which will function as the source file folder.

   For important information about file locations, see *Source File Folders*.

4. **Create exception files in**: Specify the full pathname of a folder where the Importer can create files that will contain records rejected during an import. The Importer needs full rights to the specified folder. If the folder is on another machine, use the UNC filename (\\SERVER\FOLDER) instead of a drive letter.

   Exception files can be edited and then reimported.

   For important information about file locations, see *Exception File Folders*.

5. **When import is complete**: Specify what to do with source files after the import.

   - **Delete processed source files.** Delete the source files (whether or not they were successfully imported).

   - **Move processed source files to this directory.** Move the source files to the specified folder (whether or not they were successfully imported). If the folder is on another machine, use the UNC pathname (\\SERVER\FOLDER) instead of a drive letter. The Importer needs full rights to the specified destination folder.

     For important information about file locations, see *Processed Source Files*.

6. For **Import File Type**, indicate the format of the import file:

   - **Fielded Text Files (Tagged/Delimited/XML).** Select this option if you are importing Inmagic tagged, delimited ASCII, or XML files. See *Importing Fielded Text Files*.

   - **Document Files.** Select this option if you are importing binary files (such as Microsoft Word documents) or non-fielded text files. See *Importing Documents*.

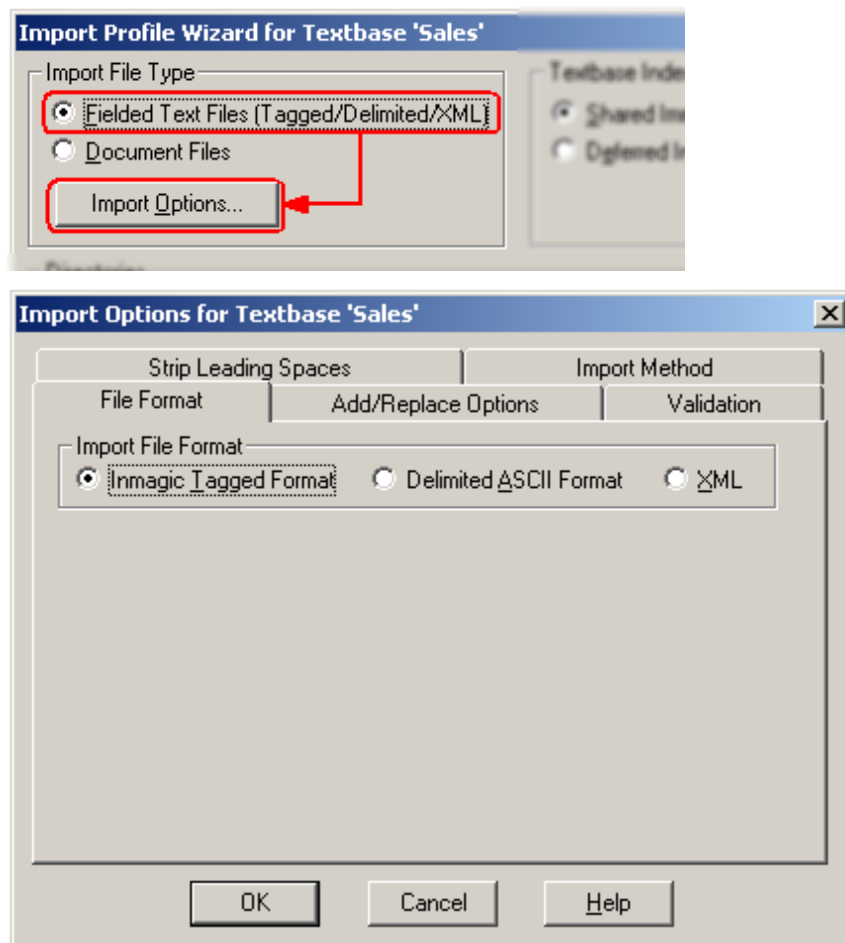7. Click **Import Options** to open the Import Options dialog box.

Your next step is to specify Import Options for Fielded Text Files or Document Files, depending on the type of file you are importing.

When you are done, close the dialog boxes until you return to the Inmagic Importer Settings dialog box, where you can click **OK**. The Importer will run at the specified interval. (If the Service is set to Manual, you need to start it first. See *Starting and Stopping the Importer Service*.) **Important!** You must restart the Service in order for your changes to take effect.

Importing Fielded Text Files

If you selected **Fielded Text Files** in the Import Profile Wizard (see *Creating an Import Profile)*, then clicked **Import Options**, the Import Options dialog box appears.

Use the Import Options to specify how to load fielded text files. Each tab is described below.

File Format tab (for fielded text files)

Specify the format of fielded text files to be imported.

Inmagic Tagged Format

Select **Inmagic Tagged Format** if your import files are in the Inmagic tagged format, the default format used for sharing files among Inmagic textbases.

Delimited ASCII Format

Select **Delimited ASCII Format** if your import files are in standard delimited ASCII format. A delimited ASCII file contains values separated by characters that indicate where records, fields, and entries end. Please see *Delimited ASCII Format (fielded text)*.

Delimiter Options

Specify the delimiters used in the file(s) you are importing. You can select delimiters from the lists, type keyboard characters, or type an ASCII value inside curly braces, for example {29}.

- **Record Separator.** The character(s) that indicate where each record ends. The default is {CR}{LF}, carriage return line feed. You could also use, for example, {LF} for a line feed only or {TAB} for a tab character.

- **Field Separator.** The character used between fields. The default field separator is a comma. Be sure to use a character that would never appear in the content of a field, unless you also specify a quote character.

- **Entry Separator.** The character(s) used between multiple (repeating) entries in a field. The default is a vertical bar (|). The following example shows three entries separated by vertical bars: Bill Jones|Sue Smith|Tom Williams.

- **Comment.** Information preceded by the specified comment character will be ignored during the import. Comment lines are often used to provide information to other people

reading the file--they are not intended to be interpreted as record information. The default comment character is an exclamation point (!).

- **Quote Character.** Specify which character is used to enclose field information that may contain delimiters. A quotation mark (") is used in this example:

```
"Jones","15 Elm Street","Smalltown","CA"
```

If a field in the import file contains a character that you have specified as a Record Separator, Field Separator, or Comment character, surround the field information with the specified Quote Character (by editing the source file). The following example uses quotation marks to cause the comma after Smith to be interpreted literally (as a comma instead of as a Field Separator):

```
"Smith, John",Fly Fishing,1987
```

Specify **{NONE}** if field information will not contain delimiters, and you want quotation marks interpreted as data (often desirable if the field separator is {TAB}).

To embed a quote character in the data without having it be interpreted as a quote character, type it twice, as was done with the double set of quotation marks in the following example:

```
"Jones (aka ""Jonesy"")","15 Elm Street","Smalltown","CA"
```

Note that Entry Separators are never interpreted literally, even if they are inside a quoted string.

Field Names for Delimited ASCII Format

If you select **Delimited ASCII Format** as the file format, select a **Field Names** option to determine how to associate information in the import file with fields in the textbase:

- **Import Fields in Textbase Structure Order.** Select this option button if the fields in the import file appear in the same order as fields in the textbase structure, and the import file contains every field that is in the textbase structure. For example, if the textbase structure consists of *Last Name*, *First Name*, and *Age*, in that order, this file will import properly:

```
Watson,Alan,41
Moselli,Robert,21
Doe,Janice,78
```

If a record in the import file contains more fields or fewer fields than the textbase structure, the record will be rejected.

- **First Row Contains Field Names.** Select this option button if the first line of the import file contains field names. This is the most flexible option, and the only option that is practical if the field order is different for different import files. Here is an example of a comma-delimited file in which the first row contains field names:

```
Last Name,First Name,Age
Watson,Alan,41
Moselli,Robert,21
Doe,Janice,78
```

If a record in the import file contains more fields than the first row, the record is rejected. If a record in the import file contains fewer fields than the first row, the fields not represented are not modified.

- **Specify Order in which to Import Fields.** Use this option if field names are not in the first row **and** all import files will have the same fields in the same order. Use this option if the import file contains fewer fields than the textbase or if the import file has fields in a different order than the textbase structure.

  When you select this option button, the Available Fields list and Import Fields list appear on the tab. Move fields from the Available Fields list to the Import Fields list. The Import Fields list should reflect the order of fields in the import file. If necessary, use the **Change Order** buttons to move fields up or down in the list. For example, if your import file looks like this:

  ```
  41,Alan,Watson
  21,Robert, Moselli
  78,Janice,Doe
  ```

  move *Age*, *First Name*, and *Last Name*, in that order, to the Import Fields list. The Import Fields list now indicates the order of fields in the import file. The file will import properly, regardless of the order of fields in your textbase structure.

  If a record in the import file contains more fields than the Import Fields list, the record is rejected. If a record in the import file contains fewer fields than the Import Fields list, the fields not represented are not modified.

XML

Select **XML** as the Import File Format if you are importing XML files. Then specify a transform to use to convert the incoming XML to the required (tagged) format.
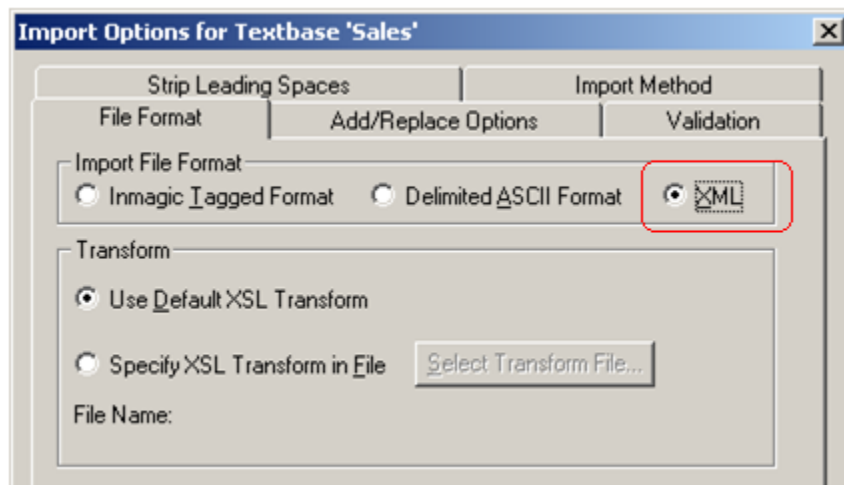
Use Default XSL Transform

The default transform uses a standard style sheet that is automatically and dynamically generated by the Importer service for use with the target textbase. You may want to use the default XSL transform in the following situations:

- If you exported XML from another textbase that has the same structure as the one into which you are importing.

- If you created an application (for example, using the XML features of *WebPublisher PRO*) that generates XML records using the structure defined in the *WebPublisher PRO* input schema.

Specify XSL Transform in File

Transforms the XML file using an XSL style sheet that you specify (for example, to make field names in the incoming records match the textbase into which you are importing them).Click the **Select Transform File** button to select a style sheet.

Note: If any records are rejected during an XML import, they will appear in the exception folder as tagged. After correcting the problems reported in the exception file, do not copy it back to the XML source folder. Instead, use a folder that accepts tagged records and uses the same matching parameters.

Add/Replace Options tab (for fielded text files)

Use the Add/Replace Options tab to specify whether to check for matching records; and if a match is found, whether to add, reject, change, or delete records during the import.
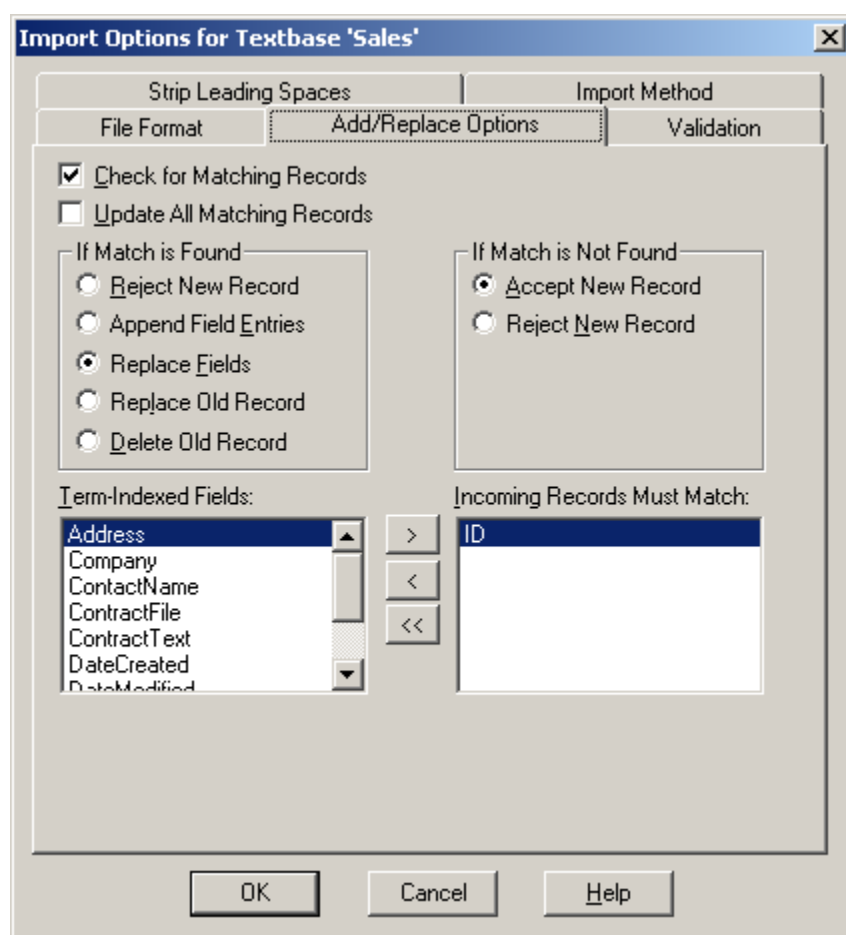
How a Match is Determined for Fielded Text Files

Matching is done during an import if you select **Check for Matching Records** on the Add/Replace Options tab.

To be considered a match, **every entry** in the matching field in the import record must match an entry in an existing record in the textbase. For best results, use a non-repeating field (one without multiple entries) that contains unique information. An Automatic Number field is ideal, because it contains unique numbers.

During an import, the program looks at the specified matching field(s) in each incoming record, and does a Term search (=) on existing records in the textbase.

If you are importing fielded text files (not documents), you can request that a match be based on more than one field. For example, if your textbase contains multiple records for the same last name (Joe Smith and Bill Smith), you might want to match the last name, first name, and phone number to ensure that the correct record is modified.

If you match on multiple fields, every entry in every matching field in the record in the import file must match an entry in an existing record in the textbase, to be considered a match. The order of fields in the Incoming Records Must Match list is not important.

Specify enough matching fields (or one unique field) to ensure a match.

Check for Matching Records

Check this box to enable the remainder of the tab so you can specify how records are added, rejected, modified, or deleted. Move at least one field from the Term-Indexed Fields list to the Incoming Records Must Match list, and choose what to do if a match is found or not found.

If you are not worried about duplicates, do not use **Check for Matching Records**. This is often the case when adding records to an empty textbase, when you are sure that the import file does not contain duplicate records.

Important! If you do not check for matching records, all of the records in the import file will be added to the textbase. No matching will be done, which may result in duplicate records. This box is cleared by default. Leave it cleared only if you are adding new records to an empty textbase or if you are sure that the records in the import file do not duplicate existing records (or other records in the import file).

Update All Matching Records

Check this box to have all matching records updated.

Clear this box to skip the update if more than one matching record is found. If more than one record in the textbase matches the record in the import file, the record will be rejected. This is the more common case.

If Match is Found

Use the options in the If Match is Found group to specify what to do if an incoming file matches an existing record.

A match is determined by comparing information in a field in an incoming record with information in that field in existing records. The field used for matching is the one specified in the Incoming Records Must Match list. For example, if the ID field is used, the ID in an incoming record is compared to the ID in existing records. If a match is found, the record is handled according to the If Match is Found option. This is explained in How a Match is Determined.

- **Reject New Record.** Select this option to add new records to the textbase, but avoid duplicate records. If an incoming record (in the source file) matches an existing record (in the textbase), the incoming record will be rejected.

- **Append Field Entries.** Select this option to merge fields in duplicate records. If an incoming record matches an existing record, all fields in the incoming record will be added as new entries to the existing record. This option adds information to a record

without changing any existing information. Field entries in the import file are appended after the last entry for that field in the textbase. Information is never appended in the field(s) used for matching (Incoming Records Must Match list). The following example matches on the *Model* field:

Existing record	Incoming record	Modified record
Model A29 Color red Size medium	Model A29 Color purple	Model A29 Color red ; purple (Color field now contains 2 entries) Size medium

- **Replace Fields.** Select this option to completely replace information in certain fields. If an incoming record matches an existing record, fields in the incoming record will replace those fields in the existing record. Fields in the existing record that are not in the incoming record remain unchanged. Information is never replaced in the field(s) used for matching. The following example matches on the *Model* field:

Existing record	Incoming record	Modified record
Model A29 Color red Size medium	Model A29 Color purple	Model A29 Color purple Size medium

- **Replace Old Record.** Select this option to replace existing records. If an incoming record matches an existing record, the existing record is deleted and replaced with the incoming record.

- **Delete Old Record.** Select this option to delete existing records from the textbase. If an incoming record matches an existing record, the existing record will be permanently deleted. Fields in the import file other than those used for matching are irrelevant and ignored. For example, an import file for deleting records might look like this:

```
ID 12
$
ID 43
$
ID 17
$
```

In this case, records with IDs 12, 43, and 17 will be permanently deleted from the textbase. If the import file contains a record that does not exist in the textbase, the record will be rejected (and written to the exception file, if you are using one). The deleted records themselves are **not** placed in the exception file.

If Match is Not Found

Use the options in the If Match is Not Found group to specify what to do if a match is not found in the textbase for a record in the import file.

- **Accept New Record.** Select this option to add records in the import file for which no match is found.

- **Reject New Record.** Select this option to reject records in the import file for which no match is found. If you are using an exception file, the rejected record will be copied to the exception file.

Term-Indexed Fields / Incoming Records Must Match

Specify a field for matching, so the Importer knows which records to modify. Select a field from the **Term-Indexed Fields** list and click the right-arrow button to move the field into the **Incoming Records Must Match** list.

For best results, use a unique, non-repeating field, such as an Automatic Number field. You can match on up to five fields at once. See How a Match is Determined.

Information in the selected field(s) will be compared during the import to determine whether a record in the import file matches a record in the textbase. For example, if the *ID* field is used, the ID of an incoming record is compared to the ID in the existing records. If a match is found, the record is handled according to the option selected in the If Match is Found group, and is either rejected or accepted into the textbase.

# Validation tab (for fielded text files)
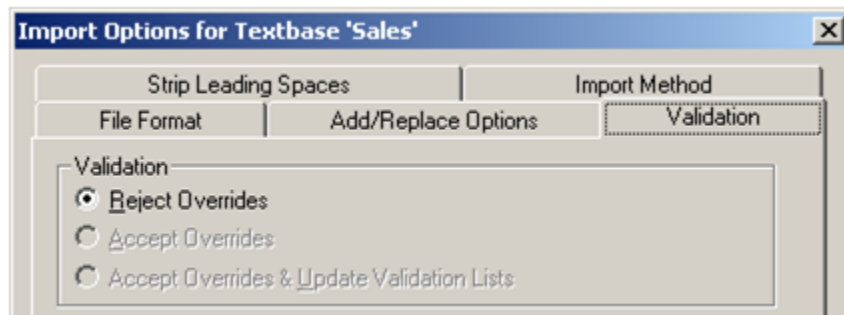
Use the Validation tab to specify how to handle records that fail to meet content validation rules (masks, ranges, and validation lists). For example, a field may have a validation mask, such as ###, that only allows digits. If an incoming record does not conform to the mask, the option selected on this tab determines whether the record will be accepted or rejected.

Rejected records will be written to an exception file, if you specified an exception file folder on the Import Profile Wizard dialog box.

- **Reject Overrides.** Records that do not meet content validation rules will be rejected during import. For example, if the *Fruit* field uses a validation list that contains only apple and orange, an incoming record that contains pear is rejected. Records that are rejected during an import are copied to an exception file if an exception file folder was specified on the Import Profile Wizard dialog box. See *Using Exception Files*.

- **Accept Overrides.** Records that do not meet content validation rules are accepted during the import. For example, if the *Fruit* field uses a validation list that contains only apple and orange, an incoming record that contains pear is accepted.

- **Accept Overrides & Update Validation Lists.** Same as **Accept Overrides**, except that validation lists are updated with entries that are not already in the list (assuming the validation for the field allows lists to be updated with override values). For example, an incoming record that contains pear is accepted, and pear is added to the validation list for that field. This provides a way to keep validation lists up-to-date, but introduces the risk of adding unwanted items to your validation lists.

Note: Overrides will be accepted only if the **User May Override Content Validation** check box is selected for the involved field(s) on the DB/TextWorks Edit Fields dialog box.

Strip Leading Spaces tab (for fielded text files)
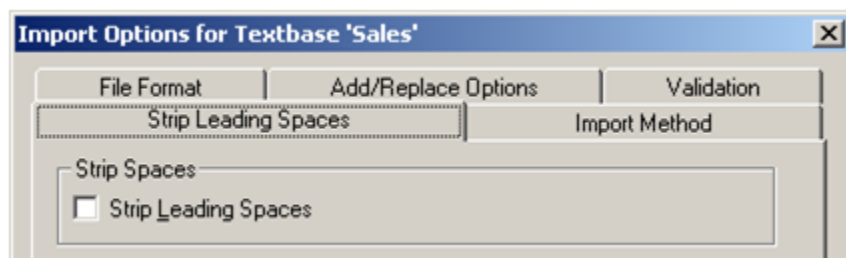
Use the Strip Leading Spaces tab to specify whether leading spaces and tabs are retained or removed from the import file during import. This will affect the appearance of imported files.

Leading spaces are spaces or tabs at the beginning of continuation lines, immediately after the field name or a new entry boundary, or after line breaks in the data.

Check **Strip Leading Spaces** if you want to remove these spaces and tabs during import.

Clear the box if you want to retain leading spaces and tabs, which may have been included for formatting purposes, such as indentation.

Note: Upon import, every tab character at the beginning of a line or embedded in the text is converted to one space. This is true regardless of whether you check or clear this box.

Import Method tab (for fielded text files)

Use this tab to select an import method. If you are not sure which method to use, click Help Me Decide.
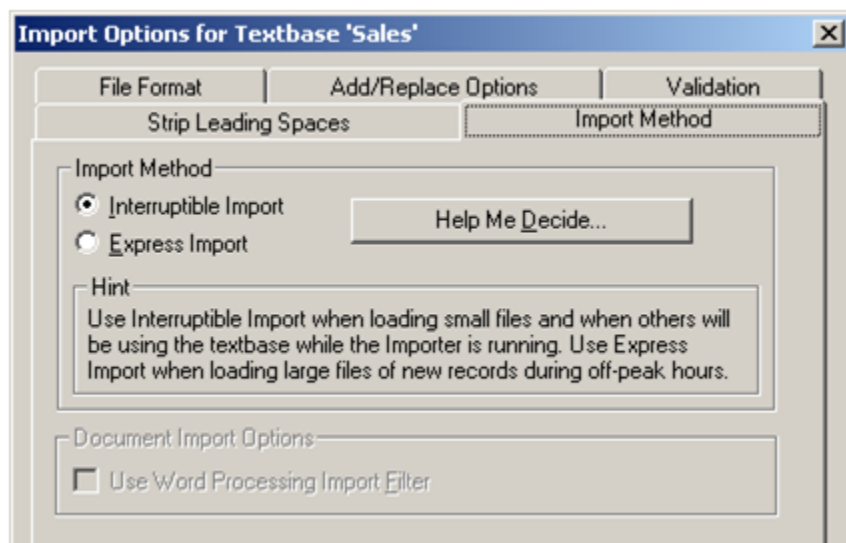
Interruptible Import

Interruptible Import is the default. Use Interruptible Import in the following situations:

- To obtain the fastest import when loading small files or files that contain just one record each.

- To retain the ability to interrupt the import quickly (for example, if you notice that many records are being rejected), including the ability to stop relatively quickly at the end of the operating interval.

- To allow others sharing the textbase to perform index operations (for example, searching and editing) during the import.

Express Import

Express Import is designed for importing large files of new records during off-peak hours. Express Import is faster when importing files that contain many new records, but may make it difficult for others to use the textbase.

For faster imports, you can specify the maximum permitted size of the Express Import buffer by editing the INMAGIC.INI file. See *Troubleshooting Imports*.

Do not use Express Import if you are modifying existing records or if you are importing files containing only one record each, because it can result in slower imports.

Express Import does most of its work in memory, keeping the indexes open until the import is done, so others sharing the textbase cannot perform record operations until the import is complete. Also, interrupting an import can be time-consuming (it may take minutes or hours before an interrupt request is processed).

Tip! If Express Import is disabled and you want to select this option, close the Import Profile Wizard for Textbase dialog box and select **Shared Immediate Indexing**.

Help Me Decide

Use the **Help Me Decide** button on the Import Method tab to decide whether to use Express or Interruptible indexing. Check the boxes below, and the wizard will recommend an indexing method. Click the **Apply Recommendation** button to use the suggested method.

- **All import files will be small (for example, a single record)**. Check this box if you are importing small files or if each file consists of a single record.

- **Import files will replace/amend/delete existing records using Add/Replace Options**. Check this box if you are checking for matching records and you are appending field entries, replacing fields, replacing old records, or deleting old records.

- **Others will be using the textbase while the Importer is running**. Check this box if you will be running the Importer during regular working hours while others are using the textbase to perform record operations, such as searching and saving.

- **There are multiple profiles for this textbase**. Check this box if the same textbase appears more than once in the Textbases to Service list.

Document Import Options - Use Word Processing Import Filter

This option is not enabled for Fielded Text File imports, because it does not apply. For an explanation of this feature, see *Importing Documents*.

Importing Documents

If you selected **Document Files** in the Import Profile Wizard (see *Creating an Import Profile)*, then clicked **Import Options**, the Import Document Options dialog box appears.
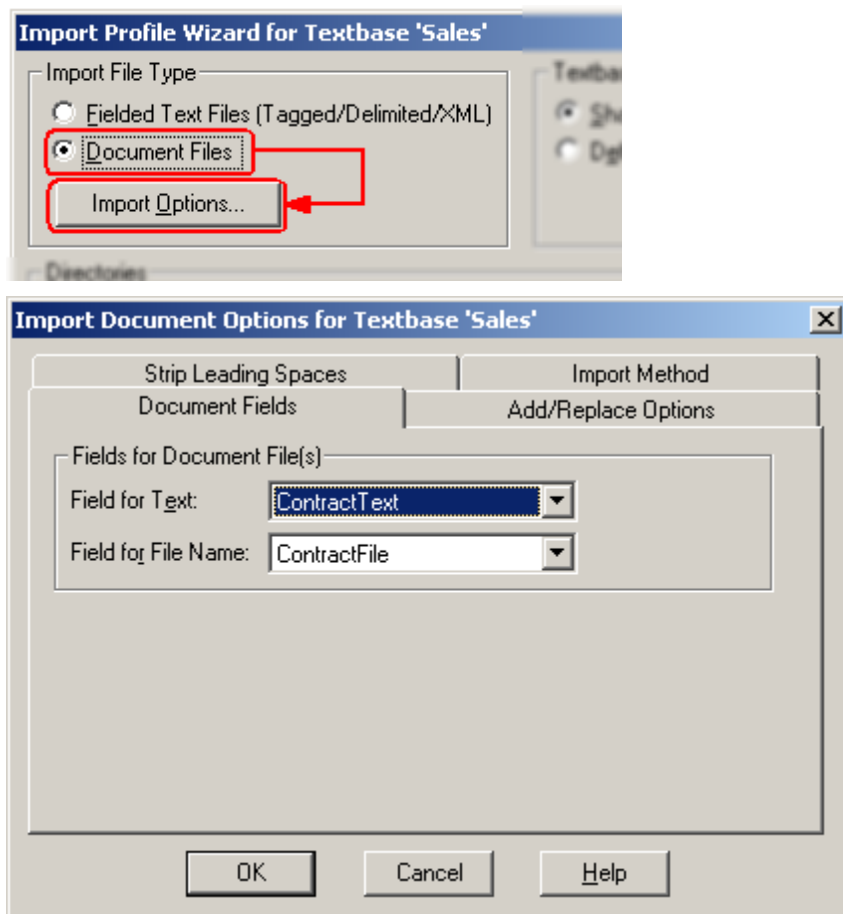
Use these options to specify how document files will be imported. Each tab is described below.

Document Fields tab

Specify the **Fields for Document Files**, explained below.

Field for Text

Select a field in the target textbase to hold incoming document text. The entire text of each incoming file will be placed in this field. Formatting is not retained, so the purpose of importing the text is to support searching, not reports.

Field for File Name

The Field for File Name will be used for matching, to determine if an incoming document matches an existing record in the textbase. Select a Field for File Name if at least one of the following is true:

- You are adding or replacing document text in existing records.

- You are creating new records and you want to avoid duplicates.

- You want to include the file name in new records.

When you do the import, the file name will be loaded into that field for new records and used for matching to modify existing records. Note the following points:

- The field you select must be a Term-indexed Text field that was created to hold the name of the document file.

- When you open the Add/Replace Options tab, the selected field will appear in the Incoming Records Must Match list. The selected field will be used for matching, if you choose Check for Matching Records.

- When a document is imported, its file name will be compared with the existing value in the Field for File Name in the target textbase. If the names match, the selected If Match Is Found operation will be performed.

- If you are modifying records, the records that you want to modify must already include a Term-indexed Text field containing only the name of the document file, *not the path*. Otherwise, the field values will not be considered a match.

Validation for Document Imports

During an import, documents that do not meet validation requirements will be rejected. You cannot override validation for document imports.

Rejected documents will be noted in the textbase log file. Exception files are not used for document imports.

Note that it would be very unusual for a field that holds document text to have validation applied. However, a field that holds the name of a document (Field for File Name) may have Unique Entries Only validation, to ensure that the filename is unique.

Add/Replace Options tab (for documents)

Use the Add/Replace Options tab to specify whether to check for matching records; and if a match is found, whether to add, reject, modify, or delete records during the import. A common use for this feature is to replace an older version of a document with a newer version. Matching on the file name finds the old version of the document. Note that file names will not match if the version in the textbase includes the path.
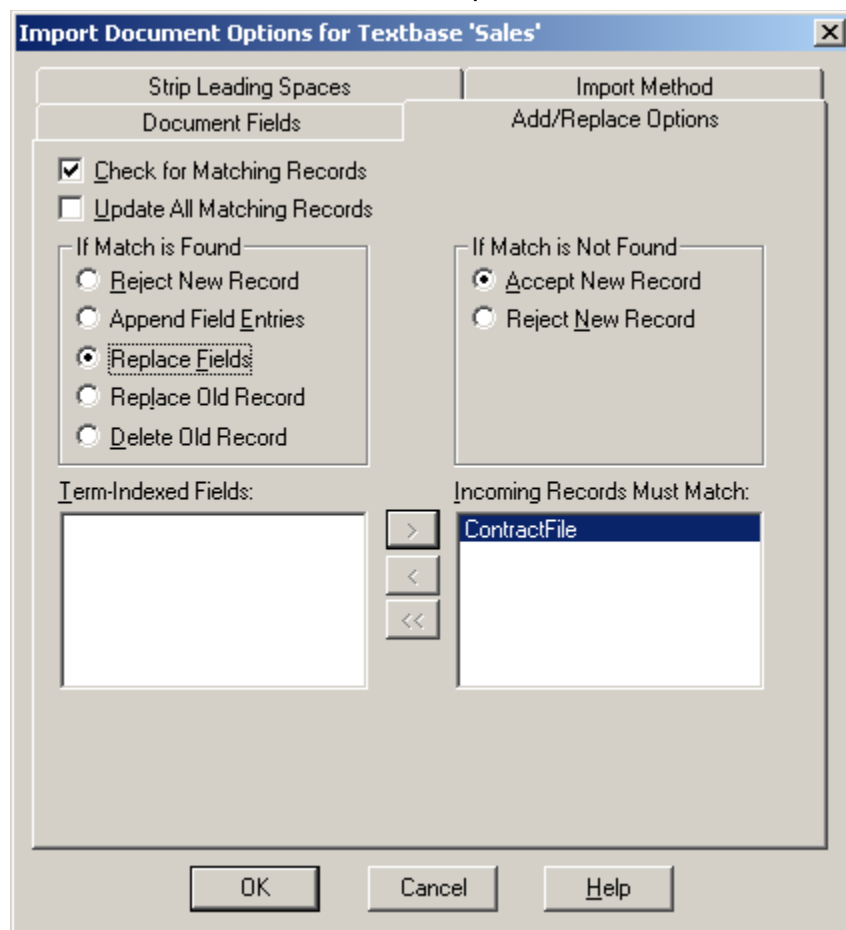
Check for Matching Records

Select **Check for Matching Records** to specify how records will be added, rejected, modified, or deleted. Then specify options to determine what to do if a match is found or not found.

Note: To check for matching records, you must choose a **Field for File Name** on the Document Fields tab. The selected field is the only field you can use for matching when importing documents. If you did not select a Field for File Name, you will not be able to check for matching records.

Important! If you do not check for matching records, the incoming document will be added to a new record in the textbase even if there is already a record containing the file with the same name. **Check for Matching Records** is cleared by default. Leave it unchecked if multiple documents coming from different sources can have the same filename (exclusive of path).

Update All Matching Records

Check this box to have all matching records updated, even if more than one record matches. This is unlikely to be the preferred outcome when importing documents. Clear this box to skip the update if more than one matching record is found.

If Match is Found

Use the options in the If Match is Found group to specify what to do with a document if a match is found in the textbase. A match is determined by comparing the document filename with the **Field for File Name** specified on the Document Fields tab.

- **Reject New Record.** Add new records to the textbase but avoid duplicates. If an incoming document has the same filename as a record in the textbase, the document will be rejected.

- **Append Field Entries.** Merge fields in duplicate records. If an incoming document matches an existing record, the document will be added as a new entry in the existing record. This option adds information to a record without changing any existing information. **Note**: This option is rarely used when importing documents, as it is unlikely that you would want to append the new version of a document to the old version.

- **Replace Fields.** Completely replace the document text. If an incoming document matches an existing record, the incoming document replaces the contents of the field selected from the **Field for Text** list (on the Document Fields tab) in the existing record. All other fields remain unchanged.

- **Replace Old Record.** Replace existing records. If an incoming document matches an existing record, the existing record is deleted and replaced with the incoming document text. If you are using matching to update old versions of documents, and you have any information in the record other than the filename and document text, select **Replace Fields** instead of **Replace Old Record**, to ensure that you retain the other fields.

- **Delete Old Record.** Delete existing records from the textbase. If an incoming document matches an existing record, the existing record will be deleted. If the filename does not exist in the textbase, the document will be rejected. **Note**: This option is rarely used when importing documents.

If Match is Not Found

Use the options in the If Match is Not Found group to specify what to do if a match is not found (in the textbase) for an incoming document.

- **Accept New Record.** If a match is not found for an incoming document, the new document will be added to the textbase.

- **Reject New Record.** If a match is not found for an incoming document, the document will be rejected during the import. Exception files are not used. For details about rejected files, see the textbase log file.

Term-Indexed Fields / Incoming Records Must Match

The field selected from the **Field for File Name** list on the Document Fields tab appears in the **Incoming Records Must Match** list. This is the only field you can use for matching. Information in the selected field will be compared during the import to determine whether a document in the

source file folder matches a record in the textbase. If a match is found, the document text is handled according to the options specified in the If Match is Found group.

Strip Leading Spaces tab (for documents)

Use the Strip Leading Spaces tab to specify whether leading whitespace will be retained or removed from the document during import. This will affect the appearance of imported files, for documents that include spaces at the beginning of continuation lines, immediately after the field name or a new entry boundary, or after line breaks in the data.
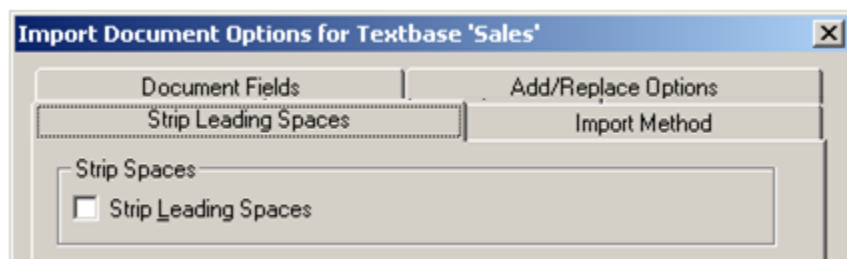
Check **Strip Leading Spaces** if you want to remove leading spaces during import. Clear the box if you want to retain leading spaces, which may have been included for formatting purposes, such as indentation. **Note**: Upon import, every tab (at the beginning of a line or embedded in the text) is converted to one space. This is true regardless of whether you check or clear this box.

Import Method tab (for documents)
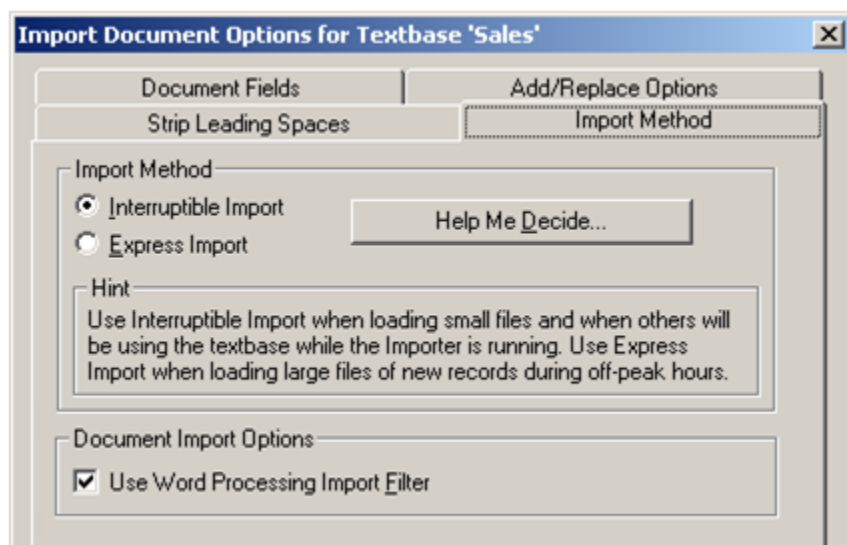
Import Method

Choose **Interruptible Import** or **Express Import**. If you are not sure which method to select, click Help Me Decide. For an explanation of the Import Method options, please see *Import Method tab (for fielded text files)*.

Document Import Options

This topic explains the **Use Word Processing Import Filter** option in the **Document Import Options** section of the Import Method tab.
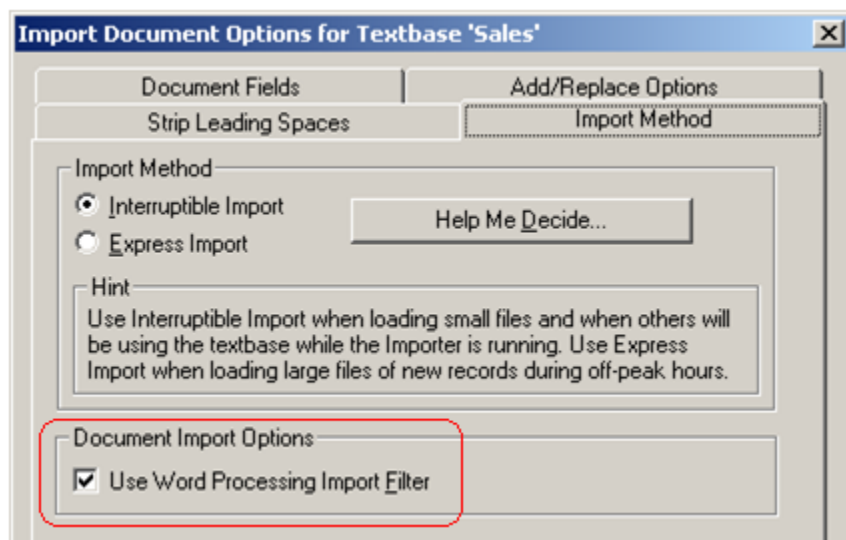
To extract text from binary (non-text) files, such as Microsoft Word documents or PDF files, IFilters must be installed on the Importer machine. Many IFilters are installed as part of the Windows operating system, but some may need to be installed manually. Please see *IFilters* in the Glossary. If IFilters are detected by the Importer, **Use Word Processing Import Filter** is checked by default.

Use the **Use Word Processing Import Filter** option in the following cases:

- If you are importing non-text files such as Microsoft Word documents or Excel spreadsheets. This ensures that the files will be passed through the import filters-- a necessary step when loading non-text files into a DB/TextWorks textbase.

- If you want to strip out HTML codes when importing HTML documents.

Do not use the **Use Word Processing Import Filter** option in the following cases:

- If you are importing text files.

- If you want to retain HTML tags when importing HTML files.

Clearing this box for other types of files (which may contain binary, non-ASCII information) is likely to cause errors during import.

You have now finished configuring the Import Options.

Close the dialog boxes until you return to the Inmagic Importer Settings dialog box, where you can click **OK**. The Importer will run at the specified interval. (If the Service is set to Manual, you need to start it first. See *Starting and Stopping the Importer Service*.) **Important!** You must restart the Service in order for your changes to take effect.

Chapter 4:   Diagnostics and Troubleshooting

This chapter provides information related to configuration and troubleshooting.

Interrupting an Import

There are several reasons why you might want to interrupt an import:

- You might notice that most of the records are being rejected, which could indicate a problem with validation or with the format of the import file.

- You might decide that you do not want to wait for the import to finish, especially if you are using Express Import, which prevents others from using the textbase during the import.

- You might decide that you specified the wrong options or file(s) to import.

To safely interrupt an import without harming the textbase, stop the Inmagic PowerPack Importer service:

1. In Windows Control Panel, open the Administrative Tools folder and double-click **Services**. Or, if you prefer, choose **Start>Run** and enter **services.msc**.

2. Right-click the Inmagic PowerPack Importer service and choose **Stop**.

You might notice a delay while the Importer finishes processing information held in memory and writes it to disk. The next time the Importer runs, it will continue from where it left off.

Power Failures and Reboots during Imports

Note: We strongly recommend that you stop the Importer before rebooting, especially if it is during the *"Run During Interval"* period. Aborting during Interruptible Import can result in one record not being completely indexed. Aborting during Express Import can result in *many* records not being completely indexed.

Never reboot the machine while the Importer is running! Rebooting often occurs after installing new software, when the application asks if you want to restart your machine.

When an import is interrupted by a power failure or reboot, the textbase may be left in a damaged state. DB/TextWorks displays an error message when you open a damaged textbase.

Damage can involve bad pointers, missing information, or both. These are serious problems that affect the integrity of your textbase. Missing index information causes inaccurate retrieval when you search. Missing records may be retrieved, but not displayed or exported.

You can continue to search a damaged textbase without making any of the problems worse. However, you should avoid performing maintenance activities (adding, editing, and so on) because of the risk of increasing the damage.

Note: Some structural problems can be repaired by running DB/Text Checker. If there is additional concern that imported information was not fully indexed, the same information can be reimported with a matching option (**Replace Old Record**). Or you can use DB/TextWorks to dump the textbase then load a new textbase.

Using Exception Files to Reload Rejected Records

An exception file is a text file that contains records that were rejected during an import. Rejected records are copied in their entirety, and each record is preceded by a comment line (starting with **!**) that explains why it was rejected.

You can edit the exception file to make corrections, then reimport it. You do not need to remove the comment lines, because they will be ignored during an import.

Example of a rejected record in an exception file:

```
! ** Entry in field 'Subject' not found in validation
list: Computeeeers
Subject Computeeeers
Title Introduction to Computers
Author Pierce, Coleman
Date 31-May-10
$
```

Each exception file has the same name as its import file, in the format FILE.x01. For example, if you import Books.txt, the exception file will be called Books.x01. If Books.x01 already exists, Books.x02 will be used, up to Books.x99.

- To specify if and where exception files should be created, use the Import Profile Wizard.

- For recommendations about where to save exception files, see *Exception File Folders*.

The format of the import file determines the format of the exception file:

- Delimited ASCII records are copied in delimited ASCII format.

- Inmagic tagged records are copied in Inmagic tagged format.

- XML records are copied in Inmagic tagged format. If there is an error converting an XML file to Inmagic tagged format, an exception file is not created.

- Documents are never written to exception files. Information about rejected documents is written to the textbase log file and/or the import log file.

To correct and reload rejected records:

1. Open the exception file in any text editor or word processor.

2. Make corrections to the records, and save the edited file as ASCII text. You do not have to remove comment lines (which start with an exclamation point) because they will be ignored during an import.

3. Move the edited file to the appropriate source file folder (as if it were a regular import file) and the Importer will load it into the textbase.

   Note: If the original file was in XML format, its exception file will be in Inmagic tagged format. Therefore, you must either create a new import profile to load Inmagic Tagged files into the target textbase, OR use DB/TextWorks to import the corrected exception file.

Initialization and Log Files

This topic describes the initialization and log files that are generated by the Importer. These files can be opened using any text editor, such as Microsoft Notepad.

InmServices.INI File

The InmServices.INI file indicates the location of all of the PowerPack Service initialization files (DBCHKSRV.INI, DBUPDSRV.INI, DBIMPSRV.INI).

To determine or change where the Service INI files are saved

1. Use a text editor to open the **InmServices.ini** file, which is in the Services installation folder (for example, C:\Program Files\Inmagic\PowerPack\InmServices.ini).

2. Locate the [InmServices] section, and change the IniFilePath value. Make sure the folder exists, as the Service won't create it. For example:

    ```
    [InmServices]
    IniFilePath=c:\InmServices
    ```

3. Save and close the file.

DBIMPSRV.INI File

The Importer initialization file (DBIMPSRV.INI) contains settings specific to the Importer service, such as the textbases to be serviced, source file folders, and whether an exception log is kept.

DBIMPSRV.INI is created by the software when you click **OK** in the Inmagic PowerPack Textbase Importer Settings dialog box. It is located in the folder specified in the InmServices.INI file (see above).

Note: Most changes to the INI file are maintained by Importer Settings user interface, but the maximum number of textbase profiles is not. To increase that number, edit the file directly using NotePad or a similar text editor.

To edit DBIMPSRV.INI

1. Close the Inmagic PowerPack Importer Settings dialog box.

2. Use a text editor to open DBIMPSRV.INI. (If you are not sure where the file is, look at the InmServices.INI file, explained above.)

3. To change the maximum number of textbases that can be added to the Textbases to Update list (the default is 50), add the TextbaseMaxCount= parameter to the [Textbases] section and specify any number up to 100.

4. Save and close the file.

5. Restart the service in order for your change to take effect.

Importer Log File

The Importer.log file provides detailed information about imports and it records service activity, such as time started and stopped.. For example, this is where information is recorded if any document files are rejected (since exception files are not used for rejected documents).

By default, the file is called Importer.LOG and is saved in the location specified in the Inmagic PowerPack Importer Settings dialog box. To specify a different name or location, click the **Specify Importer Log File** button.

Best practice is to always use a log file, call it Importer.LOG, and keep it in the same location as the DBIMPSRV.INI file.

Whenever there is activity, new information is appended to the end of the log file. You may occasionally want to delete the file, if it becomes very large. A new one will be generated the next time the Importer starts (if **Use Importer Log File** is selected).

Troubleshooting Imports

To troubleshoot Importer issues, use the following resources:

- Read the optional log file (see *Importer Log File*).

- Click the **Importer Information** button on the Inmagic PowerPack Importer Settings dialog box to display and/or print information about the Importer, including the parameters defined for each import profile. This can help you determine which profile is associated with a textbase on the list, especially if the textbase appears more than once.

If you see the message "Cannot proceed because textbase is inaccessible" it is possible that the textbase does not exist. You may also get this message if the textbase resides on a machine other than the one where the Importer resides, and:

- The account you are using cannot see the textbase. Log in using an account that is able to see all the textbases in the Textbases to Service list. For example, use the same account that was used to add the textbases to the list.

- The textbase was added using a drive letter instead of the UNC file name. Remove the textbase from the list, then add it again using its UNC file name (for example, \\SERVER\VOL1\LIBRARY\CATALOG instead of G:\LIBRARY\CATALOG).

Glossary

deferred updates
Records that were added, edited, or deleted while Deferred indexing was in effect.

delimited ASCII format
A standard format for text files, in which values are separated by characters, such as commas or tabs, which delineate records, fields, and entries. For example, a comma-delimited file uses commas to separate fields.

document files
Binary files or text files that are not fielded or tagged. Each document file will be loaded into one record. The entire plain text content of the incoming file is placed in one field.

exception file
A text file that contains records that are rejected during an import and a comment describing the reason.

Express Import
An import method intended for loading large files of new records during off-peak hours.

fielded text files
Text files in Inmagic tagged, delimited ASCII, or XML format. A fielded text file may contain one or more records. Characters within the file delineate fields, entries, records, and so forth. For example, a dollar sign ($) indicates the end of a record in an Inmagic tagged file.

IFilters
Plug-ins that allow Windows to search inside different file formats. For the Importer to extract text from documents in formats other than plain text, the workstation that initiates the import must have the appropriate IFilters installed. Many IFilters are included as part of the Windows operating system. For example, the Microsoft Office Filter (OffFilt.dll) is used for MS Office files (doc, xls, ppt, etc.). IFilters for other third-party document formats can be downloaded from the Web. Note that the Adobe PDF IFilter may be installed with Adobe Acrobat. If you do not have Acrobat installed, or it is installed but no IFilter is available, you must download and install the necessary IFilter from the Web first.

index
A collection of keys (words or terms) derived from information entered in fields. DB/TextWorks uses indexes to provide fast access to information.

Inmagic tagged format
The default format for sharing files among Inmagic textbases.

Interruptible Import
An import method intended for loading small files and/or for loading files when others will be using the textbase while the Inmagic Importer is running.

knowledge agent

An application that obtains information from a particular source, such as a database, and writes it to a supported format so it can be imported.

log files

The import log file (IMPORTER.LOG, by default) is a text file that records service activity, such as time started and stopped. The textbase log file (<textbase>.LOG) lists any document files that were rejected.

source files

Files that will be imported into a textbase.

Windows service

A program that runs in the background under Microsoft Windows and is listed in the Services Control Panel.

XML

Extensible Markup Language. Used for structured documents and data on the Web. The Importer can import data files in XML format.

Index