ALE Learnset Manager

Administration Guide

Version: 2.2

Written by: Documentation Team, R&D
Date: December 2021

Hyland

Table of Contents

ALE Learnset Manager Overview

The ALE Learnset Manager or ALM is a web-based administration client that enables you to create, modify, and delete projects and classes which will be used by the Automated Learning Engine (ALE) to learn how to classify or extract data from documents. You can also edit classes by adding and removing training documents to improve the performance of each learnset.

Access the ALE Learnset Manager

You can connect to the administration client on any server where the ALE Learnset Manager is installed. The client URL is usually `http://servername:8080/ALM/`.

Note: This URL is case-sensitive.

Unless your installation is protected by a single sign on (SSO) provider, you will be directed to a login screen when you access the ALE Learnset Manager client.

To log in, complete the following steps.

1. In the upper text field, type your username.
2. In the lower text field, type your password.
3. Click **Connect**.

Note: When the ALE Learnset Manager is protected by an SSO provider, the SSO system handles the authentication. If you are authorized to use the application, you are directed to the **ALM Viewer** module after authentication. If you are not authorized, an error message is displayed, and you are unable to use any module. ALE Learnset Manager does not enforce password complexity.

Understand the interface

The following image helps you to understand the interface of the ALE Learnset Manager web application and how to use the user interface elements to navigate.
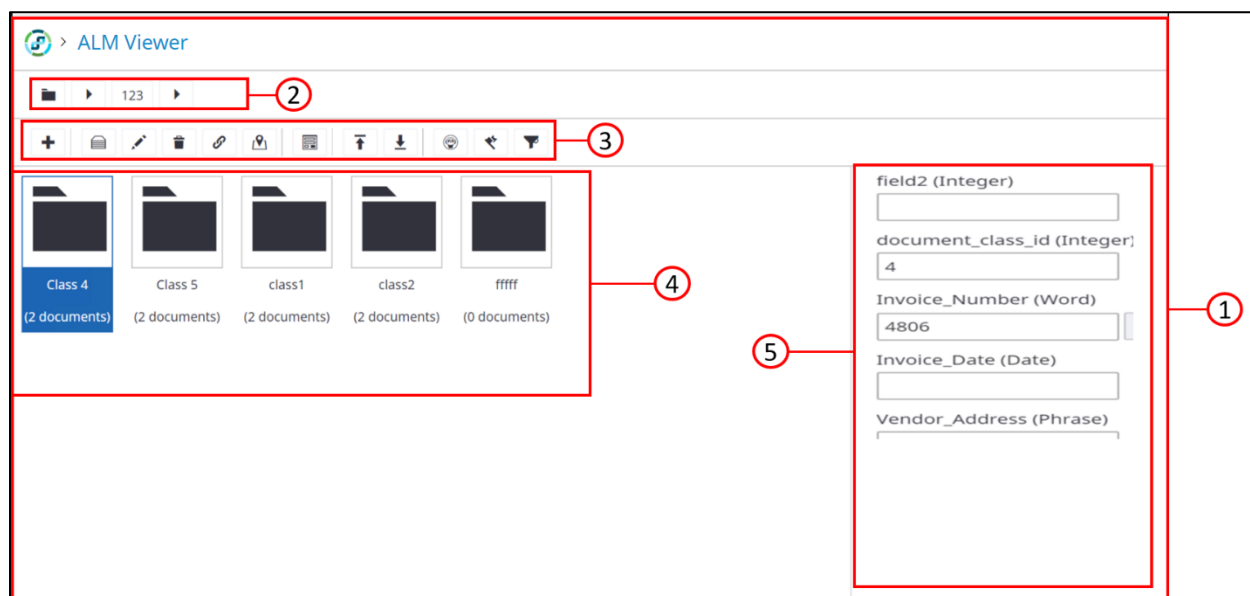
Figure 1: ALE Learnset Manager Interface

Element	Name
①	ALM Viewer
②	Project Navigator
③	Main menu
④	Viewer area
⑤	Fields viewer area

Additional Information: The element number ② referred to as the Project Navigator is the first row of icons in the ALE Learnset Manager user interface. It allows you to navigate through projects and classes, and acts like a breadcrumb to display the context.

(folder icon) displays the available projects in the viewer area. This screen is referred to as the project viewer.

(right arrow icon) displays a menu of projects or classes depending on which level you are in.

- Selecting one of the projects in the project viewer will open its class viewer and its name will appear in the project navigator.

- Selecting one of the classes in the class viewer will open the first document and the class's name will appear in the project navigator.

Terminate your session

Your session terminates automatically after a period of inactivity (typically 30 minutes). To terminate your session immediately, complete the following steps.

- On the **ALM Viewer** page, click and then . If you logged in directly to the administration client, you are redirected to the login page. If your installation is protected by an SSO provider, a message is displayed informing you that your session has ended, and you are prompted to close your browser window.

Work with User Authorization

User authorization enables you to select authenticated users who can use ALM. A user must be both authenticated and authorized before they are allowed access to ALM.

Modify the list of authorized users

To modify the list of authorized users, complete the following steps.

1. On the project navigator, click . A list of projects is displayed in the project viewer area.
2. Click .
3. In the **Edit Authorized Users** dialog box, in the **Authorized Users** field, enter the list of authorized usernames. You can use commas to separate the entries from one another.
4. Click **Save**.

Work with the Project Viewer

A project consists of field definitions, classes, and training documents.

The project viewer area displays a list of all the projects that are available to you.

Access a project

To access a project, perform one of the following tasks.

1. On the project navigator, click ▶ next to 📁 and select a project from the drop-down menu.
2. Do any of the following:

- Double click on the thumbnail of the project whose details you want to access.

- Click the thumbnail of the project whose details you want to access and then click 📑.

Create a project

To create a project, complete the following steps.

1. On the project navigator, click ➕.
2. In the **Project Name** field, enter the name of the project.

   o Select the **Use positional information for classification** checkbox, if you want positional information to be taken into consideration when learning the project,

   o Select the **Use UTF-8 encoding** checkbox if you want to pass your documents in the UTF-8 format.

3. Click **Create**.

Modify a project

To edit a project, complete the following steps.

1. On the project navigator, click 📁 to display a list of all projects.

2. Click on a project's thumbnail and then click ✏️.
3. Modify the information in the **Edit Project Settings** dialog box as required. For more information, refer to the Create a project topic.
4. Click **Save**.

Delete a project

To delete a project, complete the following steps:

1. On the project navigator, click 📁 to display a list of all projects.

2. Click on a project's thumbnail and then click 🗑️. A confirmation message is displayed.
3. Click **Yes** to delete, and **No** to go back to the project viewer.

Schedule Global Learning for a Project

This section enables you to configure a learning schedule for your projects.

Activate Global Learning Schedule

To configure all projects to learn on a set schedule complete the following steps.

1. On the main menu, click . The **Schedule Learning** dialogue box is displayed.
2. Select the **Activate** check box and enter the following information.

    a) In the **Time Interval** field, enter a value that denotes the interval in which the project is scheduled to learn.

    b) In the **Time Unit** field, enter the unit of measurement for calculating the time. For example, days, hours, or minutes.

    Note: The **Time Interval** and the **Time Unit** fields are used in conjunction with each other to create the schedule. For example, entering the **Time Interval** value as *30* and selecting *Minute* as the **Time Unit** will trigger a schedule learning every 30 minutes.

    c) [Optional] Select the **Start scheduling based on specific time of a day** check box and enter the start time in hours and minutes in the **Schedule Start Time (HH:mm)** box. This activates the learning schedule at a particular time of the day, using a 24 hours-based clock.

    Note: The start time is based on the local time of the server.

3. Click **Submit**. The project learning schedule is activated.

Modify Learning Schedule

To modify an existing learning schedule, complete the following steps.

1. On the main menu, click . The **Edit Scheduler Configuration** dialogue box is displayed.
2. Modify the learning schedule information, as required.
3. Click **Edit**. The changes are saved.

Deactivate Learning Schedule

To de-activate an existing learning schedule, complete the following steps.

1. On the main menu, click . The **Edit Scheduler Configuration** dialogue box is displayed.
2. Clear the **Activate** check box and then click **Edit**. The project learning schedule is de-activated.

Work with the Class Viewer

The class viewer area displays a list of the classes in the project. To access a class, perform one of the following steps.

- On the project navigator, click ▶ next to the project name and select a class from the drop-down menu.

- In the class viewer, double-click a class's thumbnail.

- In the class viewer, click a class's thumbnail to select it then click 🗎. The first document of the class is displayed in the viewer.

Create a class

To create a class, complete the following steps.

1. On the project viewer area, click a project's thumbnail to select it and then click 🗎.
2. Click ➕.
3. In the **Class Name** field, enter the name of the class.
4. Click **Create**.

At this point, you can only add a new training document.

Rename a class

To rename a class, complete the following steps.

1. On the class viewer area of a project, click a class's thumbnail to select it and then click ✎.
2. In the **Class Name** field, enter the name of the class.
3. Click **Rename**.

Delete a class

To delete a class, complete the following steps.

1. On the class viewer area of a project, click a class's thumbnail to select it and then click 🗑. A confirmation message is displayed.
2. Click **Yes** to delete and **No** to go back to the class viewer.

Merge two classes

To merge two classes, complete the following steps.

1. On the class viewer area of a project, click a class's thumbnail to select it and then click 🔗.
2. On the **Merge Classes** dialog box, select the class that is to receive the documents.
3. Click **Merge**.

Once the documents have been moved to the destination class, the selected class will be deleted.

Work with Project Fields

Project fields consist of the definitions of the fields that ALM can extract.

You can define the name, type and description of these fields that appear in the document viewer form. The field's type determines the values that are valid for that field. For more information on field validation, refer to the Validation chapter.

Each project has a field that contains the detected document class which by default is called **document_class_id**. This field is generated automatically. You can modify its name but cannot delete it. You can add, modify, and delete other fields.

Based on the project fields, you can define class fields that are relevant to a particular class.

View project fields

To view a project's fields, complete the following step.

- On the class viewer area of a project, click a class's thumbnail to select it and then click 🖳 . The **Edit Project Fields** dialog box appears with the names of the existing fields.

Add a project field

To add a project field, complete the following steps.

1. On the class viewer area of a project, click a class's thumbnail to select it and then click 🖳 .
2. In the **Edit Project Fields** dialog box, click ➕ .
3. Enter the following information.
    a. In the **Name** field, enter the field name as it should appear in the document viewer. **Note**: Field names cannot contain spaces.
    b. In the **Type** list, select the type which represents the values to be extracted. For more information, refer to the Validation chapter.
4. Click **Save**.

Modify a project field

To modify an existing field definition, complete the following steps.

1. On the class viewer area of a project, click a class's thumbnail to select it and then click 🖳 .
2. In the **Edit Project Fields** dialog box, click the field to be modified. The field's definition is displayed.
3. Modify the information, as required.
4. Click **Save**.

Note:

- For the field that contains the detected document class, you can modify only the name.

- If you change the field type in a project which has existing training documents, you must revisit each document and check that the entry is still valid.

Delete a project field

To delete an existing field definition, complete the following steps.

1. On the class viewer area of a project, click a class's thumbnail to select it and then click ▦ .
2. In the **Edit Project Fields** dialog box, click to select the field that you want to delete and then click 🗑 . A confirmation message is displayed.
3. Click **Yes**.

Note: You cannot delete the field that contains the detected document class.

Work with Class Fields

A project consists of several classes, and the definition of the fields that it can extract. You can define class fields to allow the number of fields displayed on the form to be reduced to only those relevant to that class. You can also replace the field name by a more relevant label and alter the order of fields.

Note: If there are no class field definitions, all project fields will be displayed for a class.

Add a class field

To add a class field, complete the following steps.

1. On the class viewer area of a project, click a class's thumbnail to select it and then click .
2. Click the checkbox next to the name of the project field that you want to use as a class field.
3. In the **Form Label** text box, enter the name of the class field as it should appear in the document viewer.
4. Click ↑ to move the field higher up the list.
5. Click ↓ to move the field further down the list.
6. Click **Save**.

Modify a class field

To modify a class field, complete the following steps.

1. On the class viewer area of a project, click a class's thumbnail to select it and then click .
2. Change the form label and position in the list as required.
3. Click **Save**.

Delete a class field

To modify a class field, complete the following steps.

1. On the class viewer area of a project, click a class's thumbnail to select it and then click .
2. Select the checkbox next to the relevant project field which you want to remove from the class field list.
3. Click **Save**.

Work with Project Learning

A project needs to be learned before it can be used for classification and data extraction. During project learning, processes like how to distinguish between the classes, and where to find the values for the declared fields are learned. The learning engine determines which fields to extract for each class by learning the project.

When you make changes to the project such as editing fields, adding, or deleting classes, and so on, you need to relearn the project before you can use it for extraction. This is indicated by a message next to the

Learn Project icon in the class viewer area.

Note:

- If you do not re-learn a project after making changes to it, the extraction will work using the previous state of the project.

Learn a project

To learn a project, complete the following step.

- On the class viewer area of a project, click . The number of classes that were learned is displayed.

  Note: To learn a project, it must have at least one field, one class, and one document with a field value defined.

Work with Export and Import

You can export a project's documents and use them in a different Columbus engine-based application or on a different machine. You can also import documents for your project.

The output of the learning process is an extractor file. You can use this file to prepare another instance for extraction without learning the project again.

A batch stream set contains all the information another instance needs to learn a project locally. It consists of two files. The PTB file contains the documents, and the CBM file contains bitmaps. Both files are exported as a ZIP archive file.

A project's documents can be exported as a ZIP archive file. The file's folder structure replicates the existing classes and includes the images, position files and value files for each document. The export produces position files with a .pos extension and value files with a .ival extension. Documents exported in this way can be imported easily with the **Import** option.

Export a project file

To export project files, complete the following steps.

1. On the class viewer area of a project, click ⬇ .
2. Select one of the following.

- **Export Stream Set Files** to download a ZIP file containing the PTB and CBM files.

- **Export Extractor File** to download the extractor file.

- **Export Documents** to download a ZIP file containing the documents in standard format.

   Note: The export fields are a union of all fields defined in all classes of the project. For example, if a field is defined at the project level but is excluded from each class, the field will not be exported from the system.

Import a project file

To import documents, complete the following steps:

1. On the class viewer area of a project, click ⬆ . The **Import Documents** dialog box appears.
2. Do any of the following:

   a) Click anywhere in the **Click to add a file, or drag and drop a file here** box. The **Open** dialog box appears.
   b) Navigate to the required folder and double-click the file, or select the file and click **Open**. The file is uploaded.

   Or,

   a) Drag and drop a file to the **Click to add a file, or drag and drop a file here** box.

   Note: When you drag the file, the cursor changes to indicate that the operation is in progress, and the **Click to add a file, or drag and drop a file here** box is highlighted. To abort the drag operation, press ESC.

3. Select one of the following options.

- **Append these documents to this project** to add the documents to the existing project

- **Replace this project with these documents** to empty this project before adding the documents.

4. Click **Import**. The documents are added to the project and a report details any errors that occurred during the import process.

   Note: The **Import** button is enabled only when a file is selected.

Work with Project Testing

You can run tests against the project file to find entries in the training documents that are probably not correct. The generated report highlights any errors that were found and allows quick navigation to the documents that are failing.

The following errors can occur.

- **Unable to find a given field value on a document**. This means that the value will not be used for learning. This is not a problem if it happens in a few training documents but if no values can be found for a field within an entire class, learning that field would fail.

  Review the document and make sure that the value is correct and is appropriate for the field type.

- Unable to extract a value for a given field when extracting the training document. This indicates that the training set is not good enough yet.

  Consider adding more training documents with further examples for this field.

- Document classified into a different class from the one that it was stored in.

  Review the document and check if it really belongs in this class. Consider removing it from the training set if you still think that the assigned class is correct - it might just not be a good example.

- As a more general issue, **classification of all documents of one class into a different class**. This usually means that the two classes are so similar that they cannot be differentiated.

  Either merge the two classes into one, or add additional training documents that emphasize the differences between the classes.

Test project file

To test a project file, complete the following steps.

1. On the class viewer area of a project, click   .
2. Click on an incident to view the training document with the errors highlighted.
3. To return to the report, click on the relevant project folder.

Work with Test Mode

You can use Test mode to allow documents from a test set to be classified and their fields to be extracted using the project definitions. A confidence level is indicated against each of the classes and values that it determines. A well-extracted document from the test set can be added to the learnset to improve the quality of the learnset. Once the project has been relearned, the effect of the additional document on the learnset can be determined by running the test documents through it again.

Create a test set

To create a test set, complete the following steps.

1. On the class viewer area of a project, click. ⊚ .
2. Select one of the following options.

* **Use Document from a File Share** to create a test set from the documents in a file share or directory that is visible on the server. In the **Create Test Set** dialog, enter a file share or directory that is visible on the server in the **File Share** field.

* **Use Uploaded Documents** to create a test set from documents that you upload from a drive that's visible to your computer. Drag the files over the **Create Test Set** dialog box and release the mouse button.
  Note: You can drag single or multiple files, or a ZIP file containing multiple files. To create a test set document, an image and a word position file with the same file prefix must be uploaded – ts1.tiff and ts1.pos, for example. This step can be repeated with more files until the test set is complete. Every time files are added, the number of *documents* that have been uploaded is updated.

* **Reuse Documents from Previous Test Set** to use the last test set that was created. If the previous test set was created from a file share, you will be taken straight to the first document in the test set; if the previous test set was created from uploaded files, the **Update Test Set** dialog will be displayed where you can, if required, add more documents to the test set.

3. Click **Test** to run through the test set documents.

Test a test set

To test the documents in the test set, complete the following steps.

1. Examine the classes suggested for the document and their confidence scores.
2. Examine the values extracted for the document's fields and their confidence scores. Click on a field to highlight the candidates in the document.
3. To add a test set document to the learnset, click **Add to Class**.

   a) Select the class to which the test set document is to be added.
      Note: The class considered to be the best candidate is preselected.
   b) Select the field values. For information on specific fields, refer to the Work with field values section.

      Note: The best candidate value is filled in.

4. Click **Save** to add the test set document to the learnset.
5. In the **Copy Test Set Document to Learnset** dialog, click **Yes** if you want to relearn the project. Relearning the project before proceeding means that the information from this document will be used when classifying and extracting other documents in the test set.
6. Navigate through the test set to view more documents and, optionally, add them to the learnset.

Work with the Document Viewer

The document viewer allows you to work with the training documents in a class.

The left pane shows the current training document; the right pane shows the training document's field values. The toolbar allows you to control which document of the learnset is being viewed and to control how the document is viewed.

Navigate through a class

When a class is selected, the first training document of that class is displayed.

Use any of the following actions to navigate through the class's training documents.

- Click ◄| to move to the first document in the class.

- Click ◄ to move to the previous document in the class.

- Click |► to move to the last document in the class.

- Click ► to move to the next document in the class.

Note: If you attempt to move to another document after you update the current one, you are prompted to save the changes before moving to the next document.

Navigate through a training document

When you select a training document, the first page displays. If there are multiple pages in a document, use the following buttons to navigate through a document.

- Click ⬅ to move to the previous page of the document.

- Click ➡ to move to the next page of the document.

Resize a page

You can change the magnification of a page to get a better overview of the page, or to focus in on its detail. Use any of the following actions to change the page magnification.

- Click [↕] to fit the height of the image to the height of the pane.

- Click ↔ to fit the width of the image to the width of the pane.

- Click ⛶ to fit the image so it makes the best use of the available space.

  Note: For portrait images, this fits the height of the page; for landscape images, this fits the width of the page.

- Click 100% to display the page at its actual size.

- Click ▬ to zoom out on a page.

- Click ✚ to zoom in on a page.

Work with images and text

You can modify the way a page's image and extracted text is displayed to assist in finding required values.

- Click [image icon] to toggle between image and text mode.

  Note: In image mode, the page's image displays, and tooltips show the extracted text as you roll over the image. In text mode, only the extracted text is displayed with its layout information preserved.

- Click [box icon] to show boxes around the extracted text.

Work with PDF and text

Documents in PDF format can only be processed if ALM is integrated with Brainware for Invoices. You can modify the way a page's extracted text is displayed to assist in finding required values.

Note: The **Show Image** [image icon] functionality does not support PDF files. Click [box icon] to show boxes around the extracted text. In text mode, only the extracted text is displayed with its layout information preserved.

Work with field values

All of the data fields defined for the project display in the pane on the right. The values extracted from the current document are filled in.

There are three ways to fill in a value for a field: by manually typing a value in the field, by clicking the document text, or by drawing a box around the region of the page that contains the text.

To enter a value manually, complete the following steps.

1. Click the field to place focus on the value entry area.
2. Type in the value.

   Note: To validate the entered value against the field type, click on another field or elsewhere on the page. For this functionality to work efficiently, it is important for a field type to be well-defined.

To enter a value by clicking on the document text, complete the following steps.

1. Click the field to put focus on the value entry area.
2. Click the text on the page. The text you selected is highlighted on the page and entered into the field.

   Note: In each case, the value is validated against the field type.

To enter a value by drawing a box around the region of the page that contains the text, complete the following steps.

1. Click the field to put focus on the value entry area.
2. Click a point on the page and, with the mouse button pressed down, move to another point on the page that encloses the text you want to select.

   Note: While the mouse button is pressed down, a bounding box displays anchored at the position of the first click and extending to the current cursor position.

3. Release the mouse button. The text within the box you selected is highlighted on the page and entered into the field.

Note: In each case, the value is validated against the field type.

4. Click **Save**.

Note: If you attempt to move to another document without saving changes to the current one, you are prompted to save the changes before leaving the document.

Validation

Each field's definition includes a type: date, amount, integer, word, or phrase. Once values are entered, they are validated against the field type.

Type	Description
Date	Many date formats are recognized including date parts separated by different characters (dot, slash, hyphen, and so on), and with month names.
Amount	Digits separated by commas and dots. There should be two digits after the last decimal point. Currency designations - $ or USD, for example – are not permitted.
Integer	Digits only.
Word	Any characters but only a single word.
Phrase	Any number of words.

Highlighting target values

All occurrences of the text in the document that match the value in the selected field can be highlighted. If the selected value appears in the document frequently, it's better to use Limited learning for that field.

Click ✏ to toggle target value highlighting on and off. Clicking on a field will highlight any matching targets within the document.

Limited learning

If a target value appears in a document several times, the system may not be able to determine which one to use for training. Limited learning allows a single target value to be selected to clarify which one of several identical values is the one to use.

To use limited learning, complete the following steps.

1. If **Target Highlighting** is disabled, enable it by clicking ✏ .
2. Select a field and click one of the highlighted targets to use for limited learning. The selected target will be used for learning. The target selected for limited learning will show a red border.

To remove limited learning from a field, complete the following steps.

1. Click ⊙ next to the field to display the context menu.
2. Select **Remove Limited Learning** from the menu.
3. Click **Save**. The limited learning is removed from the field.

Work with Training Documents

Training documents are added to classes to provide additional examples for a learnset.

Note: In order to learn a project, all training and test documents must be uploaded with word position files.

Each training document must include one or more image files and a word position file. The image file can be either a single TIFF file (containing one or more pages), or one or more PNG files. If you are using PNG files for a training document with multiple pages, the PNG file names should have the page number as a suffix. For example, invoice.pos, invoice_1.png, invoice_2.png would be the files required for a two-page document. The word position file defines the text and location of each word detected in the image file and is typically derived from the output of Optical Character Recognition (OCR) software.

The availability of the word position file is shown in the Viewer's status bar.

You can view a training document for which there is no word position file and manually enter and save field values but you cannot click on a word or use rubber banding to add words to the field values as this requires the word position file to exist.

The word position file is a comma-separated text file where a line represents a single word in the image. Each line uses the following format.

- Word id

- Page number (page numbering starts at zero for the first page)

- Top left corner – x position in pixels

- Top left corner – y position in pixels

- Width in pixels

- Height in pixels

- The word

Example: 0,0,453,345,273,89, Software

Optionally, a file giving the values for some or all of the project's fields can be provided. This field value file is a tab-separated text file where a line represents the information for a single field. Each line uses the following format.

Example:

Field Name	Field Type	Value
total	amount	1,533.74

Add training documents

To add a training document, complete the following steps.

1. Select the class.

2. Click ⬆ . The **Upload** dialog box is displayed.
3. Do any of the following:

   a) Click anywhere in the **Click to add a file, or drag and drop a file here** box. The **Open** dialog box appears.
   b) Navigate to the required folder and double-click the file, or click the file and click **Open.** The file is uploaded.

   Or,

a) Drag and drop a file to the **Click to add a file, or drag and drop a file here** box.

Note: When you drag the file, the cursor changes to indicate that the operation is in progress, and the **Click to add a file, or drag and drop a file here** box is highlighted. To abort the drag operation, press ESC.

- Click **Create** after all the required files are selected. The training document is added to the class and it is selected and ready for you to edit the field values.

  Note:

- The **Create** button is enabled after image and word position files are selected. A message at the top of the dialog box changes to indicate what files still need to be added.

- To remove a file from the **Upload** dialog box, click 🗑 next to that file.

Delete training documents

To delete a training document, complete the following steps.

1. Select the class and navigate to the document.
2. Click 🗑 in the toolbar.
3. Click **Yes** to delete the document, or click **No** to abort the delete process. If you click **Yes**, the training document is deleted and the next one displays.

   Note: The class is also deleted if it contains no more training documents.

Move training documents

To move a training document, complete the following steps.

1. Select the class and navigate to the document.

2. Click ↗ in the toolbar.

3. In the **Move Document** dialog box, select the class that is to receive the document.

4. Click **Move**.