Perceptive Intelligent Capture Verifier

User Guide

Version: 5.6.x

Written by: Product Documentation, R&D

Date: Monday, November 10, 2014

perceptivesoftware
from Lexmark

Table of Contents

What are Verifier and Advanced Verifier?

Perceptive Intelligent Capture is a product suite designed for automatically processing incoming documents. Intelligent Capture can process documents from arbitrary physical sources and paper-based documents as well as from electronic files, e-mail, or faxes.

What happens when Intelligent Capture Verifier processes documents?

Structured or unstructured document input is obtained by scanning paper-based documents from fax servers, e-mail servers or as files. All documents are stored on a computer's hard drive. Intelligent Capture monitors specified directories on this hard drive for new documents. If new documents are detected, Intelligent Capture imports them.

Imported documents are first analyzed to determine the document layout and to recognize structures such as words, lines, logos, or tables.

The documents are then classified according to predefined categories. Examples of typical categories used in classification are invoices, orders, offers, or resumes. Categories can be defined individually, depending on the needs of your organization. Using a set of sample documents, Intelligent Capture actually learns to tell which category a previously unknown document belongs.

For each category, the data relevant for further processing is different. For example, if you are processing invoices, you probably want to know the total sum to be paid. This information is irrelevant if you are processing resumes, where the applicant's name, the desired position, and the contact options are more important. Intelligent Capture identifies and extracts data that is relevant for the respective document category. The data that is to be extracted can be defined individually to suit the needs of your organization.

Finally, the documents, their category assignments, and the extracted information are released from Intelligent Capture and written to designated export directories. The documents are then forwarded to connected systems. For example, invoices can automatically be forwarded to the software system used in your company's accounting department, while resumes are sent to Human Resources.

All this is done without human intervention once the Intelligent Capture application has been set up. But what happens if Intelligent Capture cannot properly process a document? There are several reasons this could happen.

- Paper-based documents might be "dirty" so that Intelligent Capture is not able to read them.
- There might be stamps or notes on the documents that make important sections illegible for Intelligent Capture.
- Intelligent Capture may encounter a document from an unknown category. Since the software was not previously trained to recognize documents from this category, it will not be able to process the document.
- Intelligent Capture may have been told to extract information that is missing, such as a form that was not filled in correctly.

That's where Verifier comes in: Verifier is the quality assurance utility of the Intelligent Capture suite. The application detects all documents with processing problems and presents them to the operator for verification.

Since the verification step is done before the export step, only qualified output will leave the Intelligent Capture process. Therefore, subsequent systems will only receive appropriate input.

The Intelligent Capture database platform for Intelligent Capture applications enables you to keep a central store of your project and authentication information. This solution also allows for central management of storage and backup and thus provides for easier security, better connectivity of your applications, and higher flexibility for your personnel.

This version of Intelligent Capture also brings an advanced graphical user interface.

These are some key features of Verifier.

- Allows central Intelligent Capture database storage of projects and of user authentication information for more flexible access.
- Allows convenient correction of automatic classification results.
- Allows convenient correction of automatic extraction results.
- Allows manual indexing of documents.
- Allows semi-automatic indexing of documents by means of database lookups.
- Allows a final check of corrected documents before release.

These are some highlights of Verifier functionality.

- The structured user interface makes the application easy to learn.
- Sophisticated status management and filter techniques show you only the documents you have to check and nothing else.
- During the application design, the user interface can be configured, providing optimum display options for each document category.
- Keyboard shortcuts are available for most operations, enabling you to get your job done quickly.
- Through automatic locking, document batches can safely be processed by teams of operators.

Verifier vs. Advanced Verifier

Advanced Verifier is an extension of Verifier that provides access to Supervised Learning – the interactive verification and training of Learnsets. Advanced Verifier requires separate licensing and may not be available at your organization.

If you've worked with Intelligent Capture before you already know that you can work with both data and text fields. Intelligent Capture with Supervised Learning and its Advanced Verifier component dramatically improve the way you work with table data. In earlier versions, the analysis of table data was "rules-based," meaning that users could not improve upon the rules established for the analysis. Intelligent Capture with Supervised Learning changes that by enabling you to interact with (or supervise) how Intelligent Capture works with your data.

Some helpful terms

This section provides some helpful terms when using Web Verifier.

Batch

A batch is just a stack of documents. Usually, this stack is not sorted. In the context of Intelligent Capture, batches consist of electronic documents. The documents inside such a batch may be paper-based documents that have been scanned to transform them into a digital format, or files created using applications such as a word processor. Various documents are normally assigned to the same batch

only because they have been received within the same time period. For example, all letters received in the morning may be scanned until noon and therefore end up in the same batch.

Folder

In a business environment, folders are normally used to keep several documents together. Intelligent Capture does the same thing with folders. However, in the context of Intelligent Capture, a folder is always a structure inside a batch. This means that batches can either consist of document stacks, or they can consist of stacks of folders.

Document

A document is a piece of information that can serve as evidence of an event, situation, or business transaction. For example, a packing slip may provide evidence that an order has actually been shipped. Since people are used to working with paper, electronic documents strongly resemble paper-based documents. You will notice that Intelligent Capture documents consist of one or several pages, though the concept of a page is not really required for digital documents.

Classification

Classification means taking an unsorted stack of documents and organizing them into smaller stacks so that each stack contains only documents belonging to the same category. In other words, you start with a mess and end up with an organized stack of invoices, a second stack of resumes, a third stack of orders, and so on. Class and category is the same thing.

Indexing

Imagine you have a homogeneous stack of invoices, and you start to write out the information contained in the documents. For each document in the stack, you will note the name of the supplier, the total sum to be paid, and the invoice number. This procedure is called indexing, and the information that was noted is the indexing information. Once you are finished, you file the invoices and use the indexing information to build your filing structure. Later, you will be able to search and identify the document with the help of the indexing information. In the context of Intelligent Capture, indexing information is applied to a set of fields associated with the document. For each document category, a different set of fields can be used.

Extraction

If you take the stack of invoices and again write out the name of the supplier, the total sum to be paid, and the invoice number, but this time automatically, the procedure is called extraction. Extraction is a means for automatic document indexing. Extraction is context-sensitive; that is, the extracted information depends on the document category.

State

A state is a number that tells you how far the processing of a document has progressed. If the entire procedure of document processing consists of single steps, then the state increases with each step that has been completed. The state also indicates whether a step has been completed successfully, or whether there have been problems. In Intelligent Capture, states are determined hierarchically from the bottom up: if anything is wrong with a document, then there is also something wrong with the batch it belongs to.

Verification

Verification is a task related to quality assurance. It involves taking a document that has been processed or partially processed, checking the processing results, and correcting any errors.

Validation

Validation is another task related to quality assurance. Validation means confirming that a processing result is correct. This can be done at several levels: for the class or a field associated to a document, for the document as a whole, or for an entire batch.

Learnset

In classification, a Learnset is a group of documents whose classification is specified by a user. For each view and each class, the user must provide a sufficient number of representative documents. Similarly, in extraction, a Learnset is a set of documents whose field contents are selected by the user from a set of candidates.

Quality assurance with Intelligent Capture

To properly ensure the quality of automatically processed documents, there are two things you need to understand.

Batches are the basic entity.

Intelligent Capture works on batches. Tasks consist of processing steps that must always be completed for an entire batch before the next task can start.

For example, if Batches 9, 10, and 11 are waiting to be classified, the application first classifies all documents in Batch 9. If this is done, the state of Batch 9 is incremented. The next task may be to classify all documents in Batch 10, or it may be to extract data from all documents in Batch 11.

What the application will not do is to classify some documents from Batch 9, then some documents from batch 10, and then go back to Batch 9 to classify the remaining documents.

If batches are the basic entities, then entire batches need to be verified and approved before they are routed to subsequent systems where other users or processes work with them.

A batch is only valid if all of its parts are valid.

An Intelligent Capture batch is valid only if all documents and processing results associated with the batch are valid. Because we are dealing with information and data, we do not use the terms “working” or “damaged.” Instead, we use the terms “valid” or “invalid.”

Intelligent Capture batches consist of a restricted number of “parts” with well-defined relationships. Therefore, we can easily see why a batch can be invalid.

A batch is invalid if:

One or more folders inside the batch are invalid.

A folder is invalid if:

One or more documents inside the folder are invalid.

A document is invalid if:

It has been classified automatically, but the classification result is invalid, or data has been extracted automatically from it, but at least one or more fields are invalid.

A classification result is invalid if:

No matching class could be found, or the class has been changed manually and not yet validated.

A field is invalid if:

The field could not be filled, the field content does not comply with validation rules that have been defined, or the field content has been changed manually and not yet validated.

Field validation rules may be violated for a number of reasons:

- The set of allowed characters may be restricted.
- Only uppercase characters may be allowed.
- There may be restrictions on the number of characters the field can contain.
- Intelligent Capture may enforce that characters which could not be certainly identified during the OCR process must be checked. These questionable results are indicated in red and are underlined.

Besides these formal validation rules, all kinds of custom rules are possible. For example, if the contents of Field 3 do not equal the sum of Field 1 and Field 2, then Field 3 may be invalid. Such a rule will typically be applied for invoices.

The application normally tells you why a field is invalid.

About Intelligent Capture workflow

In Intelligent Capture, the flow of incoming documents follows a sequence of standard processing steps. Some steps can be skipped, but the order of steps is fixed.

Automatic steps are executed by the Runtime Server and include document import with batch creation, OCR and layout analysis, classification, extraction, export, and clean-up. These automatic steps are completed with two manual verification steps that ensure only high-quality output is produced: verification of the classification and extraction steps.

If the Runtime Server has completed an automatic step and the batch contains only valid results, the next automatic step can be accomplished without human intervention.

However, if the Runtime Server detects that the batch contains invalid results, the batch must be routed to a verification station, where you can analyze and resolve the problem using Verifier. Invalid batches are presented to you in a task list, called Batch View. You will have to resolve each problem and validate each correction before you can release the batch. Subsequent automatic steps can be carried out only after release. Finally, when Intelligent Capture has finished processing a batch, the documents are sent to their recipients.

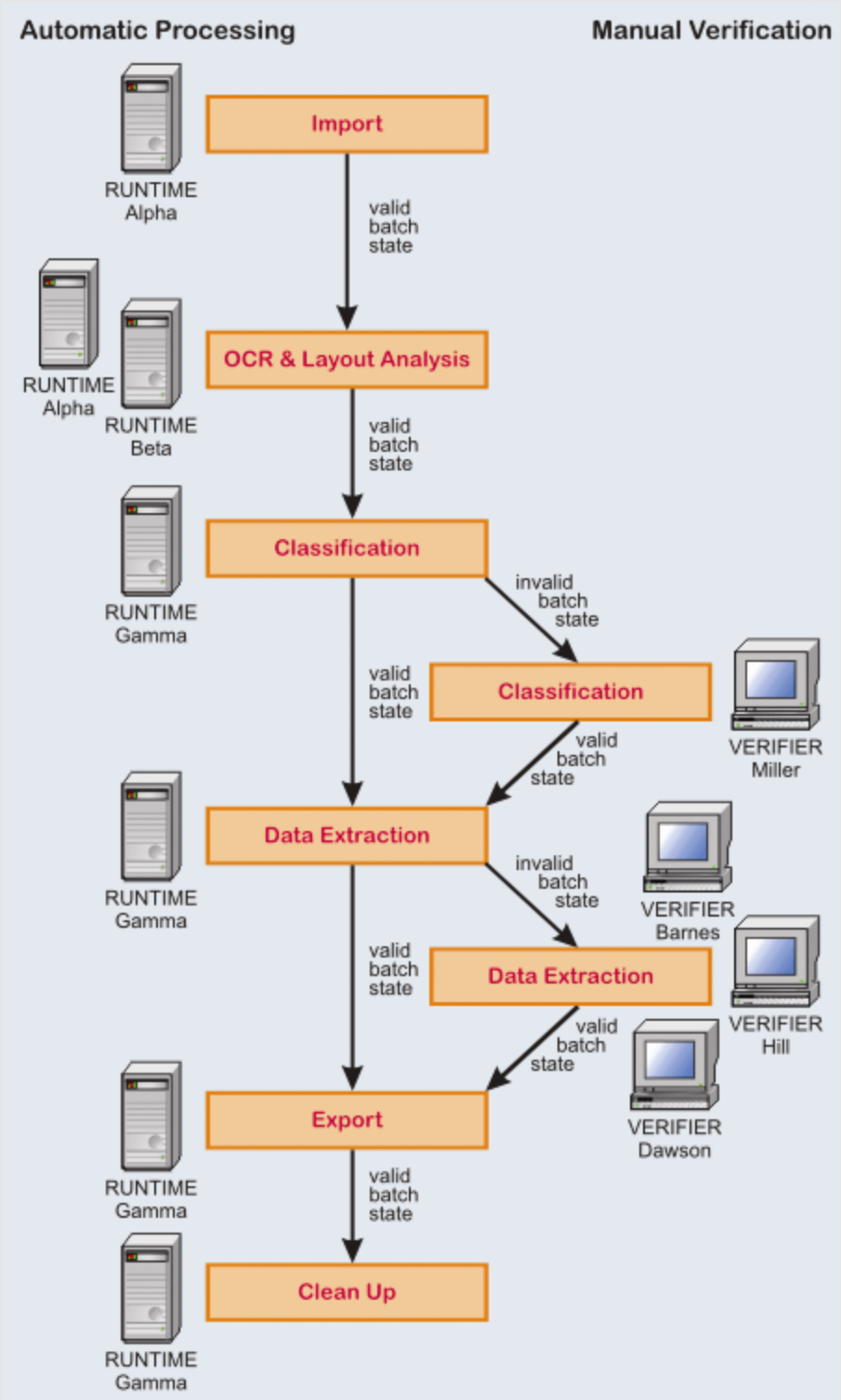
One of the objectives of Intelligent Capture is to get documents to their recipients as quickly as possible. On the machine side, automatic steps can be distributed to several computers to ensure that no delays occur. They can simultaneously perform the same or different tasks. Similarly, on the human side, Intelligent Capture supports a variety of task distributions in a team. For example, there can be specialized workstations, where one station's operator is only in charge of classification results and the other station's

operator verifies the extraction results. This can be realized by configuring Verifier accordingly. In addition, several operators can carry out the same task at the same time, but on different batches. This is possible through a locking mechanism that avoids conflicting results by making sure that a batch cannot be changed by several persons at the same time.

As a practical example, consider two servers with Runtime Server called Runtime_Alpha and Runtime_Beta which share the time-consuming task of OCR. A third server Runtime_Gamma is in charge of the remaining automatic steps. If Runtime_Gamma generates an invalid classification result, the corresponding batch is routed to a member of the QA team named Miller. Miller corrects these results using the Verifier instance running on Verifier_Miller.

Normally, correcting invalid extraction results requires more effort than correcting invalid classification results. Therefore, three other members of the QA team — Barnes, Hill, and Dawson — share a common work list containing batches with invalid extraction results. If either of them starts processing a batch, this batch will be locked for the others. Verifier sets a corresponding marker in the work list.

The following sample workflow showing how processing steps can be distributed. This configuration involves several servers running Runtime Server, and a number of workstations with Verifier that are used by a quality assurance team.


Automatic Processing
Manual Verification
Import
RUNTIME Alpha
valid batch state
OCR & Layout Analysis
RUNTIME Alpha
RUNTIME Beta
valid batch state
Classification
RUNTIME Gamma
invalid batch state
valid batch state
Classification
VERIFIER Miller
valid batch state
Data Extraction
RUNTIME Gamma
invalid batch state
VERIFIER Barnes
valid batch state
Data Extraction
VERIFIER Hill
valid batch state
Export
VERIFIER Dawson
RUNTIME Gamma
valid batch state
Clean Up
RUNTIME Gamma

Start and exit Verifier

About starting Verifier

If Verifier was installed as recommended by the setup program, you can launch it from the Windows Start menu. After startup and login, the application displays the Batch View.

Note that the Sax Basic scripting engine is no longer supported. If your application is configured to open a project which uses this engine for custom scripts, the following error message displays.

```
Project XXX.sdp is using an unsupported Sax Basic engine and cannot be used.
Please open this project in the Designer application to upgrade to the newer
engine for custom scripting (WinWrap Basic).
```

Click OK to close Verifier. You need to open the project in the Designer application and migrate it to use the WinWrap scripting engine.

About logging into Verifier

When you log in to an existing project in Verifier, you must supply your user name and password. This password is not the same as the one you use to log in to your workstation. Instead, it is specific to Verifier, and possibly to the project. However, you probably have the same user name and password for all Verifier projects you work on. Your user name and password were assigned to you in Designer when your project administrator configured the project.

Your user name and password enable you to get into Verifier and govern what you can do once there. If you have questions or problems with your user name or password, contact your project administrator. If you forget your password, your administrator can reset it for you.

Your project administrator can give you the option to remember your user name and password between logons. This has been enabled if the **Remember password checkbox** appears on the logon form. To remember your user name and password between logons, fill in your user name and password and select **Remember password** before clicking OK. Next time when you logon to the same computer, the system will fill in the user name and password automatically so that simply clicking OK will log you in.

When launching Verifier for the first time, the application is not yet configured. The Batch View will be empty and an error message displays. Verifier needs to be configured. This should be done by an experienced user.

About specifying login information with command line arguments

To suppress project authentication when starting Verifier, you can specify logon information as command line arguments. The command line argument for user name is “/USR” and for password it is “/PWD”.

For example, the following line in a Windows batch file placed in the Intelligent Capture program folder launches Verifier under John Smith’s account.

```
start /B DstVer.exe /USR "John Smith" /PWD john1234567
```

You can use the same mechanism from the Windows Run menu.

```
"C:\Program Files\Perceptive\Perceptive Intelligent
Capture\bin\DstVer.exe" /USR "John Smith" /PWD john1234567
```

If the password is empty, there is no need to specify the “/PWD” option. For example:

```
start /B DstVer.exe /USR "Guest User"
```

The administrator can also review who has logged into the application by entering a certain script. Refer to the Scripting Guide for more information.

Exit Verifier

To quit Verifier, complete the following step.

- On the **File** menu, select **Exit**.

Users, groups, and roles

To load a Verifier project, you must log in with your user name and password. If you work on more than one Verifier project, your user name and password are probably the same for all of them.

The user name/password combination not only lets you in a project, it governs what you can do once you get there. User name/password combinations were set up by your project administrator while configuring your project in Designer.

Your administrator also set up user groups and assigned you to at least one of these groups – perhaps more than one. In turn, the administrator assigned one or more roles to each user group.

There are six roles: Administrator, Learnset Manager, Supervised Learning Verifier, Verifier, Verifier Settings and Verifier Filtering.

Administrator

The Administrator role is to manage users, groups, and user-to-group assignments. Administrators install the system, configure applications, and manage data. They also design and maintain projects. This role is the most powerful of the six roles, because it encompasses the permissions for all other roles.

Learnset Manager

The Learnset Manager role is to define, modify, and maintain the Learnset.

Supervised Learning Verifier

The Supervised Learning Verifier role is to collect and manage local training data. Supervised Learning Verifiers are subject-matter experts who can propose Learnset candidates to improve system performance.

Verifier

The role of the Verifier group is to verify documents that the system could not automatically process.

Verifier Settings

The role of the Verifier Settings group is to allow changes to the Verifier configuration.

Verifier Filtering

The FLT role is to allow a Verifier user to configure custom filtering of batches. By application design, FLT users would be able to use the filtering feature even if they do not have the SET role.

Change your password

To change your password, complete the following steps.

1. From the **Options** menu, select **Change Password**.
2. Type in your existing password.
3. Enter the new password.
4. Re-enter the new password to confirm that you typed it correctly.
5. Optional. Select **Decode password**, which allows you to see the password you typed in when changing password.
   - If **Decode Password** is unchecked, when you type the password, it is masked with asterisks.
   - If **Decode Password** is checked, when you type the password, it shows the letters you type.
6. Optional. Select **Update password in Database**, which is only available if you have enabled the **Allow Database Authentication** option on the **Users** tab in the Designer **Authentication Settings**.
7. Click **Apply** and then click **OK**.

About configuring Verifier and Advanced Verifier

You can only change the Verifier settings if you have been assigned the Verifier Settings role.

Configuring Verifier entails specifying which batches of documents are processed at a given station. This includes the following.

- Sourcing of the batches either from the file system or from the Intelligent Capture database.
- The location of the batches in the file system.
- The Designer project file that contains the settings used to process the documents.
- The processing steps that you want to verify: classification, extraction, or both.
- The status of batches before and after processing.

It also entails configuring 508 Compliance, but this is done at the workstation level, not the project level.

After you configure a project's settings, you can load and save them using commands on the File menu. When loading or saving a project you can load or save a file with or without network data. When loading, click on the file type drop down box and select either <project name> (*.sdp), or <project name> skip learn data (*.sdp).

Note: You can only work with Verifier after these settings are established. Only experienced users should change the settings.

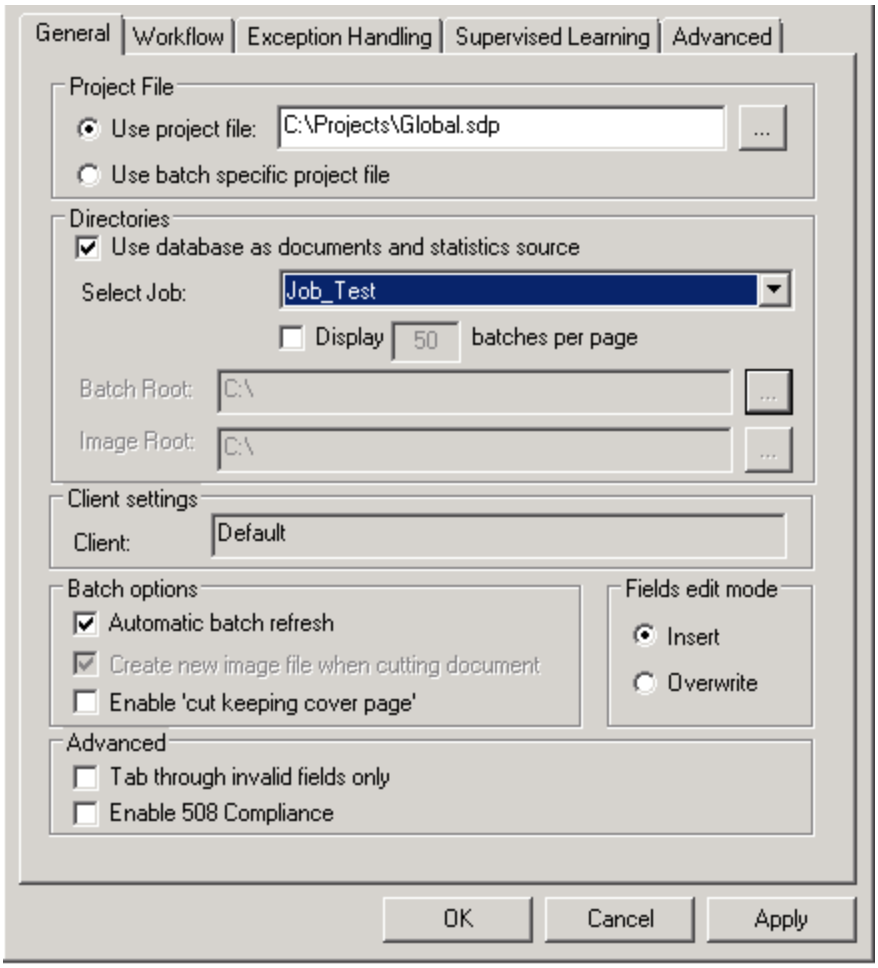
Configure Verifier

To configure Verifier, complete one of the following options.

- Select **Settings** from the **Options** menu.
- Click the **Settings** button on the toolbar.

About the General settings

The General tab is the place for general settings. It allows you to configure your referenced directories and files. Also, you can choose the Intelligent Capture database as your document and statistics source here.



About specifying the project file

The **Use project file** option is used to select the path and file name of the Intelligent Capture application that processes the documents and which contains the design of the indexing windows that you use to verify the extraction.

Select this option if you are working with a copy of the project file that resides on the local machine's hard drive, or if you use a project file from the network drive but with a mapping that is different from Runtime Server's drive allocation.

Refer to the following scenarios.

- Let's assume that Runtime Server runs on a computer called system_srv and uses the project file SampleCorp.sdp located on system_srv in C:\Perceptive\Projects\Sample. Corporation Intelligent Capture runs on a second computer called system_qa. To access the SampleCorp.sdp project file on system_srv, system_qa needs to be able to access the directory with the project file over the network. This is done by mapping the remote directory on system_qa. The mapped directory will then obtain a new drive letter.
- Let's assume that C:\Perceptive\ on system_srv is mapped to drive E:\ on system_qa. This means that C:\Perceptive\Projects\SampleCorporation\SampleCorp.sdp on systemsrv and E:\Projects\SampleCorporation\SampleCorp.sdp on system_qa are the same file.

If you use this option and your documents are processed using more than one project file, you must change the settings every time you change the project file.

The **Use batch specific project file** option uses the Intelligent Capture project file that is referenced in the batch itself. Select this option if Verifier and Runtime Server run on the same machine.

You can also select this option if they run on different machines, but in this case you need to make sure that Verifier and Runtime Server use identical paths and file names for all project files.

About specifying the directories

With the **Use Database as Document and Statistics Source** option, Intelligent Capture core information can be stored in the Intelligent Capture database. Furthermore, you are able to select the job you want from the Select Job list if you have selected the Database as your source.

Note: The file system functionality is still supported.

The **Display … batches per page** option enables you to set the number of batches displayed per page. It allows values from 1 to 200 batches to be displayed per page. This is only available if Use Database as documents and statistics source is selected. The value of 50 is set by default.

The **Batch Root** directory is where the batch control files are located. This directory also includes a license directory where you can find the network license file. When multiple instances of Runtime Server and Verifier are running they will check this directory for the network license file. If you use different Batch Root directories simultaneously, you can set up a specific path for the network license. However, you can only configure this path manually for Runtime Server and Verifier. This option is not available when using the Intelligent Capture database.

Note: When using Database, Batch Root directory licensing cannot be used, and a license share is required.

The **Image Root** directory is where subdirectories with the scanned images can be found. As a rule, Batch Root and Image Root should be the same. In special cases, for security reasons for example, the Image Root can be different from the Batch Root.

About specifying client settings

The **Client** option refers to the intent to use client-specific variables. Currently, only the default setting is available. In Designer, project administrators can define global variables for different clients. With the default entry, global variables do not vary by client.

About specifying the batch options

If the **Automatic Batch Refresh** option is checked, the Batch View automatically shows newly generated batches with matching states. If you do not want the automatic update you can clear the checkbox. This leaves you the option to refresh the Batch View (using the refresh option on the View menu) when you need up-to-date information.

The **Create New Image File When Cutting Document** option enables Verifier users to create new TIFFs when a Workdoc is split into multiples. The TIFFs correspond to the new Workdocs.

Note: This feature is disabled when the Use Database option is enabled. Any project that wants to use this feature is recommended to use the File System instead of Database.

The **Enable Cut keeping cover page** option enables Verifier users to cut a long document, such as a multi-page fax, into several shorter documents while still retaining the cover page of the original Workdoc as the cover page for each of the newly created shorter documents. If this is checked, the shortcut menu in document browsing view has additional menu entries. The new documents must then be re-OCR'd.

About fields edit mode

Use the **Insert** option when a document is opened that requires correction or confirmation of extraction results, the cursor is automatically placed in the first invalid field. If you select Insert mode, the cursor is inserted to the left of the field contents.

Use the **Overwrite** option when a document is opened that requires correction or confirmation of extraction results, the cursor automatically appears in the first invalid field. If you select Overwrite mode, the entire field content is selected.

About specifying tabbing behavior

If you select the Tab through invalid fields only option, when a user presses TAB, SHIFT + TAB, CTRL + TAB, or CTRL + SHIFT + TAB to tab through the fields in Document Verification mode, the system tabs through invalid fields only. When the user presses TAB inside of a table control, the system tabs through invalid table cells only.

Enable 508 Compliance

The **Enable 508 Compliance** option activates 508 Compliance settings for your workstation. The Enable 508 Compliance option enables 508 Compliance for all projects you work with from this station. Users at other workstations who do not want to use these features do not have to use them, even if they work on the same projects you do. To enable the 508 Compliance options, complete the following steps.

1. On the **General** tab, under **Advanced**, select the **Enable 508 Compliance** checkbox.
2. Refer to the following features available when 508 Compliance is enabled.
   - A blue arrow shows which field has focus.
   - Additional visual indicators besides color highlighting help distinguish between invalid fields, valid fields, and questionable fields. These indicators are present in table fields and form fields. Green check marks show valid fields, red Xs show invalid fields, and orange question marks show questionable fields. Field candidates are highlighted in yellow, but do not have additional validity icons.

- All menu items have underscored letters available by `ALT` menu shortcuts.
- Pop-up menus for workflow state lists and exception handling can be activated by the right-click key on the keyboard. This key is on the right of the standard keyboard, in between the Windows key and the `CTRL` key.
- In **Show Selected Batch**, the right-click keyboard key activates the shortcut menu for **Append this document to previous one** and **Cut pages into a new document**.
- During document verification, pressing `CTRL+M` or selecting **Show Selection Context Menu** activates the shortcut menu for the currently selected item.
- In the highlight columns for interactive learning mode, unmapped column items are indicated by a blue rectangle without icons while valid / invalid column items are indicated by rectangles with a “valid” / “invalid” icon at the left side of every item.
- If input focus is lost for any reason, the user can manually restore it from the Main Menu (by selecting Restore Focus) or by pressing `CTRL+N`.

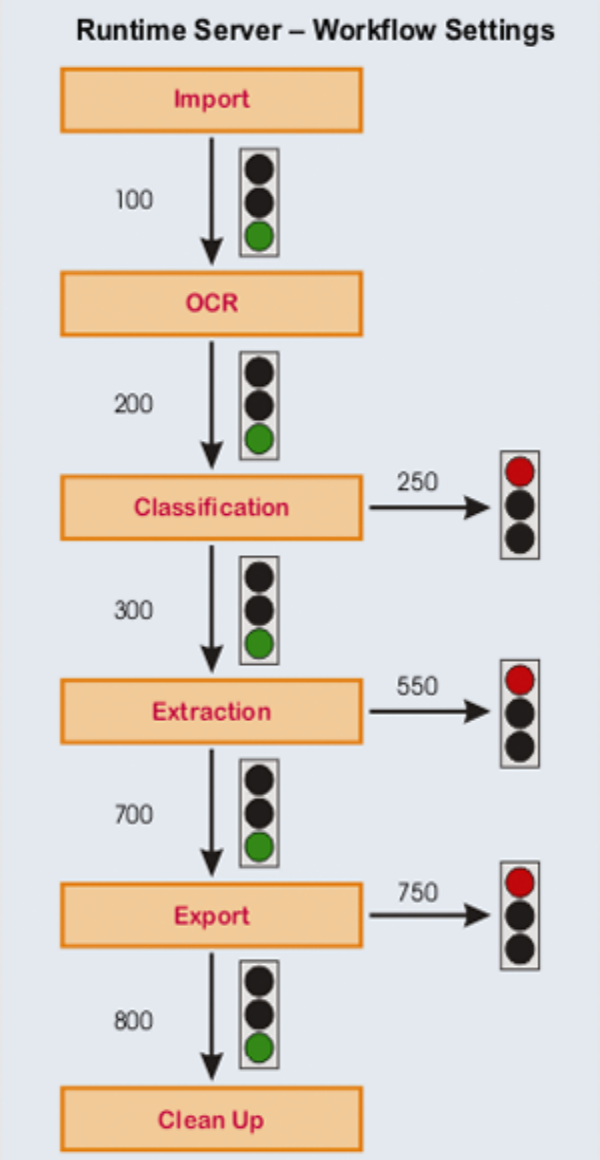
About the workflow settings

Runtime Server assigns pre-defined output states to batches after each processing step. Different states are used to distinguish successful steps from failures. If the state indicates a failure of the latest processing step, or if a step is to be carried out manually, the corresponding batch is forwarded to Verifier, where you correct the errors and supply missing results. Therefore, you need to know the output states used by your Runtime Server installation. They determine the input states used by the Verifier stations.

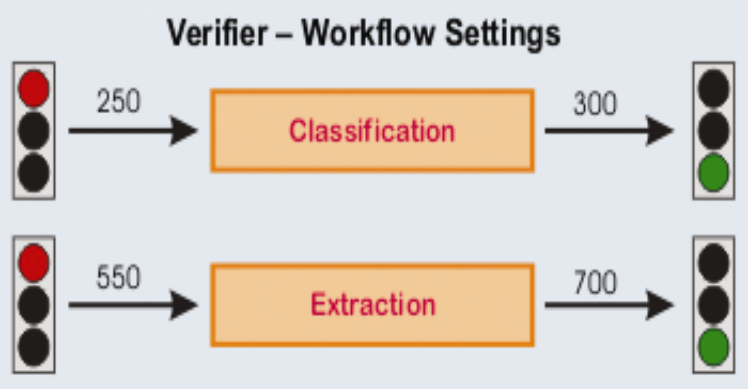
Note: If 508 Compliance is enabled, a blue arrow indicates the current focus for each field when the TAB key is used to move between the fields. The input batch status fields are not highlighted when focus is applied using the TAB key.

After verification, each batch must be returned to Runtime Server. Again, the output state of the verification step and the input state of the following step must match.

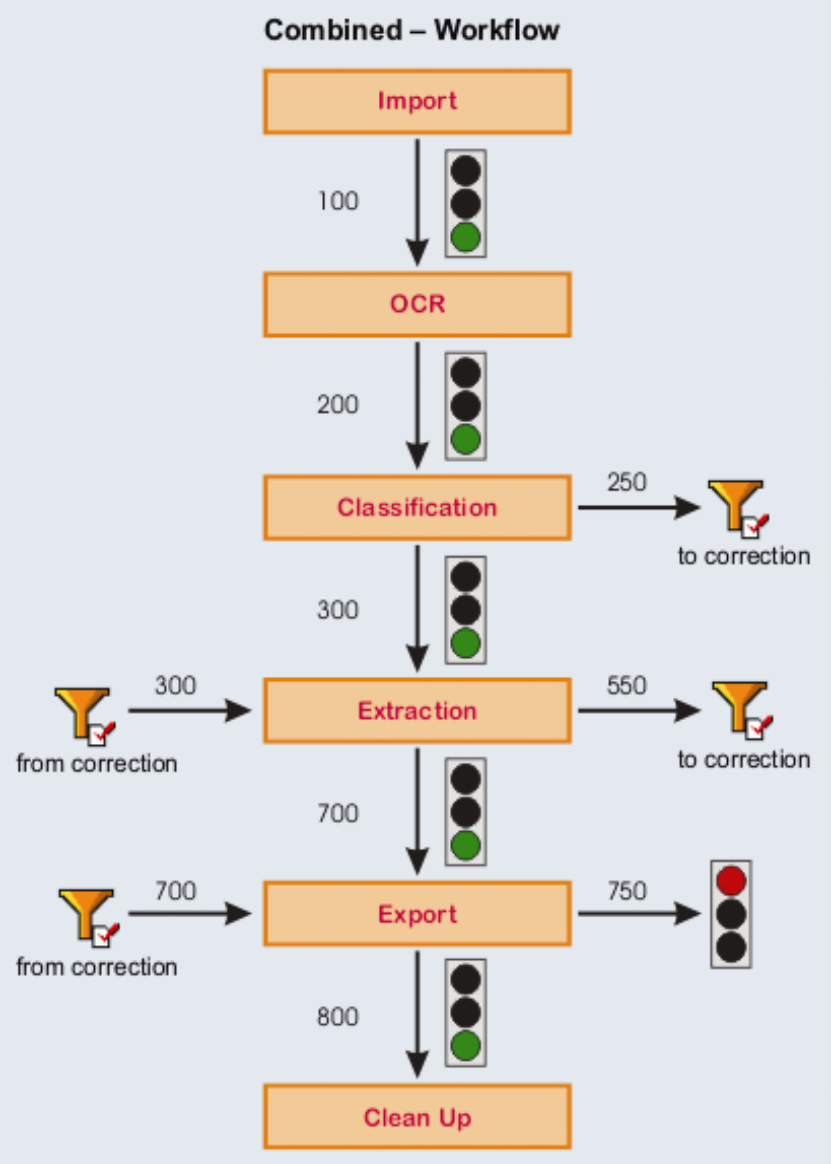
For example, let’s assume that Runtime Server uses the input and output states depicted below to process batches.



In this case, a Verifier station conducting classification, extraction verification or document separation should use the settings shown to verify whether this is correct.



A Verifier workflow configuration matching the Runtime Server settings above produces the following combined workflow.



Normally, the workflow should be configured with the following settings.

- Manual indexing / extraction correction after manual classification. Note that if a document could not be classified, automatic extraction probably will not work.
- Export after manual indexing / extraction correction.

However, there is no need to have everything done at one workstation. Tasks can be distributed among multiple workstations.

The export step is normally defined in custom scripts. Therefore, there is no default mechanism to handle export failures. However, there is the option of implementing a custom routine for this purpose. This should be done by an experienced user.

Configure tasks to perform at the workstation

To specify the tasks that are to be carried out at the current Verifier station, complete the following step.

1. Select the **Workflow** tab.
2. Complete one or more of the following options.
   - To configure document separation, click **Document Separation**.
   - To configure classification verification at this workstation, click **Classification Verification**.
   - To configure extraction verification, click **Extraction Verification**.

   Note: These steps can be performed at the workstation.

Configure the input and output states

After you have selected the steps to perform, establish values for input and output states. To add an input value, complete the following steps.

1. Right-click on the **Input** option box and select **Add State** on the shortcut menu.

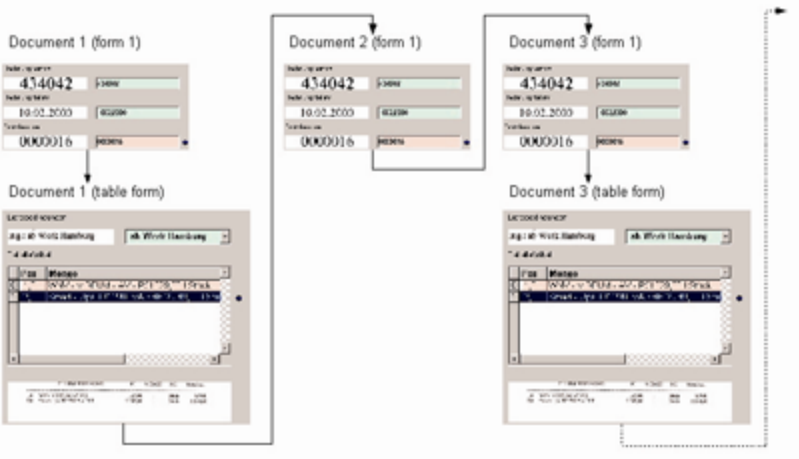
   Note: You can also change states and delete states this way.

2. To set an output value, select the value from the drop-down box to the right of the classification or extraction button.

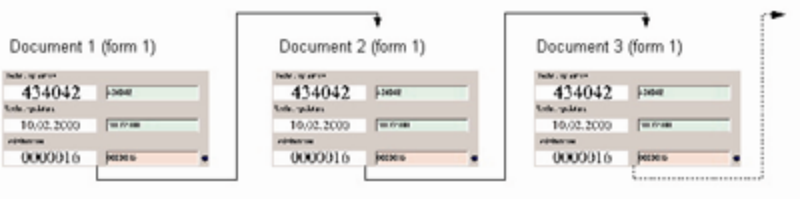
About verification rules

Verify document for the lowest input verification state only verifies a document using the lowest input verification state. When this option is selected, the correction of the documents is grouped. After the verification of each input state, the user is asked to release the batch even if there are still documents with a higher input state left to be corrected. This option is valuable when you use several forms to verify extraction fields. If you have several forms defined for default processing (meaning that this option is not selected) all forms will be shown for the document that is corrected.

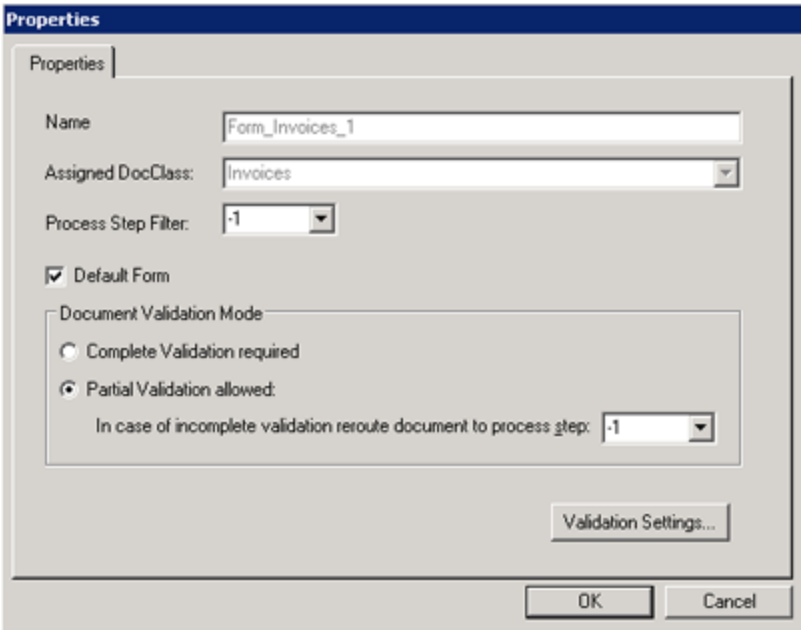
In the example shown below, first form 1 and afterwards form 2 for Document 1 is shown, afterwards form 1 for Document 2 that has only Form 1. Then all forms for Document 3 will be shown.



If this option is selected, the processing can be changed so that Form 1 for Document 1 is shown first, followed by Form 1 for Document 2 and Form 1 for Document 3 until all documents of the lowest input level are processed. Then the documents for the next input state and other forms are displayed for correction, as in the below example.



Use the Designer Verifier Design Mode to design verification forms and to define which form is shown for which input state, as in the below example.



In the settings of the first Verifier form of this document class, Partial Validation is selected, and the value is set (in this example the value is 210) which will be the output status for the document when verification with this form is completed. When a document has the input state 210, the second verification form is used to validate the fields. Therefore, the Process Step Filter was set to 210.

- **By default open the first available invalid batch and not the selected one**: By default, the first available invalid batch is opened and not the selected one. The first invalid batch is selected based on Priority (higher first), user's custom filter, sort settings and the Batch ID. This is for projects with large amount of batches and simultaneous Verifier users. It decreases time delay of project verification. This option is selected by default. And under Verification Mode, option "Verify first invalid batch" is also selected by default. If the user changed the option from the Verification Mode, this option will only remain valid for the current application session. As soon as the user quits the application, the default behavior is re-initialized through the "By default open the first available invalid batch and not the selected one" option.

- **Perform automatic extraction after classifying documents manually**: Forces Intelligent Capture to attempt to automatically extract data after the Verifier operator manually classifies the document. To select this option, the output state of the Classification Verification workflow step must be entered as an input state for the Extraction Verification input step.
- **Keep showing current document after saving**: Displays the current document after performing a Save, instead of automatically displaying the next document.
- **Allow immediate copying of selected area to a field or table cell**: Allows copying of a selected area to a field or table cell when verifying. Speeds up the process by copying single words and candidates to verification elements.

About the exception handling settings

To specify what to do if the verification cannot be finished normally, select the Exception Handling tab.

A document with an unexpected error may not be suitable for Verification. Without a mechanism to handle unexpected failures, operators will not be able to remove the batch with this document from their task list. This is why Intelligent Capture Verifier applications incorporate an exception handling mechanism. It allows operators to manually assign special states to documents with unexpected errors.

The corresponding documents can be forwarded to verification stations that specialize in collecting exceptions.

Select states

For each selected state, a menu command is available in the Verification View. The menu commands allow for case-specific handling of various types of unforeseeable errors. The description represents the menu command's label. To select a state and set its description label, complete the following steps.

1. On the **Exception Handling** tab, to select a state, mark the corresponding checkbox.

   Note: The available exception states cover the range from 601 to 699. A batch state corresponds to the lowest document state within the batch. Routing batches using their exception state is only possible if the state for successful verification is greater than the one used for exceptions.

2. To set the description label, right-click on the existing label and select **New description**.
3. Type the label into the corresponding field and confirm.

   Note: The **New Description** field allows 128 characters.

Configuring exception handling

The following settings are available for exception handling.

Before moving a document to an exception state, save it automatically

Saves a document automatically before moving it to an exception state. This applies only to the respective current document.

Create new batches with documents marked for exception handling

When this option is selected, the documents that are marked for exception handling will be moved to an exception batch. A batch is created for each exception code. The new batch receives a new batch ID. Documents from all verified batches are moved to the same exception batch in the Batch View. These batches can be released manually or automatically.

When this option is turned off, documents marked for exception handling stay in their batches. These batches keep their batch ID but are renamed according to the state description.

Automatically release all available pending exception batches that contain N or more documents or older than M minutes

When this option is selected, an exception batch is released once it contains more than N documents or is older than M minutes. This allows critical exception documents to be processed without waiting for manual intervention. Exception batches will also be released when user has exited the application and logged in again.

By default, set exception mode to Batch

If this option is selected, the scope of the command Move to Exception State on the Options menu will change to Batch automatically.

Allow user selection of exception mode (Batch vs. Document)

Enables the dynamic changing of the exception mode on the Options menu (document verification view). By default, this option is switched on. Use this option together with the option above to preserve a specific exception mode for the different user groups.

Configure Supervised Learning

The Supervised Learning tab is not available unless Supervised Learning was enabled for the project in Designer. To configure supervised learning in Verifier, with an Intelligent Capture database, complete the following steps.

1. Open the **Settings** dialog from the **Verifier Options** menu > **Settings**, or clicking the **Settings** icon and selecting the **General** tab.
2. Select the **Project** file, **Batch Root**, and **Image Root path** along with the **Database** checkbox.

   Note: Project and Batch root need to be configured with a database job in RTS.
3. Set the **Input** states for various workflow stages appropriately in the **Workflow** tab.
4. Click the **Supervised Learning** tab and select the option for **Activate Supervised Learning Workflow**.
5. Select the **Local Project** from the Verifier workstation directory.

   Note: The local project must be configured with its own base directory (Local Learnset) and Batch root.
6. Select the **Knowledge Base** directory (Common Learnset folder).
7. Select the option **Use Database as knowledge source**.

   Note: To use the option **Use database as knowledge source**, you need to have a job for a Common Learnset in the database. If this Common Learnset job is not available, a database job must be created from RTS with Common Learnset folder as the Batch Root.
8. Select the option **Distribute Local Learnset to Knowledge base** only if you want to push the Local Learnset to Common Learnset. This is required to verify the documents in Learnset Manager.

9. Select the option **Apply local classification and extraction automatically**.

   Now the settings can be saved, and the SLW user should be able to perform classification and automatic extraction in Verifier. An SLM user can now launch Learnset Manager to verify the Learnsets from the Common Learnset in the Accumulated Documents Browsing mode.

Supervised Learning settings

Supervised Learning is the interactive verification and training of Learnsets. Selecting the Activate Supervised Learning Workflow checkbox enables Supervised Learning.

The following table provides a list of the Base Settings for Supervised Learning.

Note: This information should be inherited from the settings your project administrator established in Designer.

Setting	Description
Local Project Name	The file and pathname for the local project.
Knowledge Base Directory	The file and pathname of the Common Learnset. The Common Learnset will be updated whenever the Local Learnset is migrated to it.
Distribute Local Learnset to the knowledge base	Automatically adds any documents added to the Local Learnset into a queue for the Learnset Manager to review if the documents are appropriate to be added to the Global Learnset. The knowledge base is often referred to as a queue of accumulated documents or Common Learnset pending review by the Learnset Manager for improvements into the project file.
Nominate for the Learnset but never train locally	This option enables you to prevent the Learnset from being trained locally.
Use database as knowledge source	Here you are able to select the desired job from the list. This list shows all batch jobs in the database, when the Use Database option is selected.
Always show state of all field locations after opening a document	Not available in this version.
Apply local classification and extraction automatically	New classes will be created using the supplier's name. A Learnset should also be created if you select this setting. When no local project or Learnset are used, Global Project and Global Learnset will be used instead.
Prompt if script forces or rejects insertion to Learnset	Will notify you if there is a discrepancy between script and your commands regarding the population of the Learnset.

The following table provides a list of Put document to Local Learnset settings.

Setting	Description
Only if adding activated by a user	If selected, a document will be added to a Learnset when a user requests it. The system will display a dialog box to confirm that the document should be added.
Only if adding activated by a user	If cleared, a document will be added to the Learnset only if the user requests it. This will be done automatically with no confirmation.
Automatically if more than N% invalid fields with Always prompt before adding	If selected, documents will automatically be added to the Learnset if the threshold you set is exceeded. The system will display a dialog box to confirm that the document should be added.
Automatically if more than N% invalid fields	If cleared, documents will automatically be added to the Learnset if the threshold you set is exceeded. No confirmation will occur.
Always prompt before adding	This option is not available.

The following table provides a list of the Learn new documents settings.

Setting	Description
Only by user request	Learning is initiated only when a user asks for it.
Before batch closing	Learning is initiated for every batch in the project each time any batch is closed.
Immediately	Learning is initiated anytime a document is added to the Learnset.

Advanced settings

Use the **Advanced** tab for additional features.

The following table provides a list of the Project file updating settings.

Setting	Description
Activate project file updating	Select this checkbox to activate the Project File updating feature.
Source project file location	The file and pathname of the source project file.

About batch filtering

The Batch Filter function enables you to specify filter conditions on which batches should be displayed. This is useful if you want to find a subset of batches in a huge job or to limit Verifier user activities.

The filtering dialog window is accessible outside the settings dialog so that users without a SET role but with the FLT role are able to filter batches. Only users having the FLT role assigned will be able to configure filter conditions.

The saved filtering settings apply to the current batch view, and the application saves them for next sessions.

Configure batch filter conditions

To configure batch filter conditions, complete the following steps.

1. Select **Filtering** from the list under **Options**, or click on the **Batch filter** icon.
2. Double-click on an entry in the left pane to select a batch attribute, and then double-click on a filter condition in the right pane.

   Example A filter condition of [Priority] >=1 specifies to display batches with Priority equal or greater than 1.
3. Use the following formats to specify the filter conditions.
   - Batch ID -> string, numeric
   - State -> numeric
   - Priority -> numeric
   - Name -> string
   - Folders -> numeric
   - Documents -> numeric
   - Client -> string
   - Last User -> string
   - Last Module -> string
   - Last Access -> date

   All string related operations (LIKE, '%', and so on) can be used with [Batch ID] value. Use numeric queries only with numeric Batch IDs. When batches were imported from the file system, then their IDs have a string format. If the batches were imported only with a RTS import instance and the IDs were not changed through script, then the IDs are numeric
4. Click **Clear Condition** to clear the filter condition setting.

Getting familiar with the user interface

Access Batch View

The first window displayed after starting Verifier is called the Batch View because it shows a list of batches. This is your work list.

- To access the Batch View, click the Batch View icon.

  Note: If Verifier is not yet configured, the list of batches will appear empty.

Batch view keyboard shortcuts

The following shortcuts are available in the Batch View.

Keyboard Shortcut	Command
CTRL + 1	Batch List View
CTRL + 2	Verification Mode
CTRL + 3	Document Separation Mode
CTRL + N	Restore Focus
F5	Refresh
CTRL + E	Release Exception Batches

Batch view toolbar buttons

The toolbar provides quick access to some frequently used commands.

Button	Description
	Display a property sheet where you can configure Verifier.
	Display a dialog box where you can configure the batch filtering conditions properties.
	If you click on the arrow to the right of this button, the available filters for the list of batches are displayed. You can select one of the following options. • All batches • Batches to Verify, Classification Only • Batches to Verify, Indexing Only • Batches to Verify

Button	Description
	Start the verification of the currently selected batch. Depending on the batch state, the batch is either displayed in the classification window or in the indexing window. From the list you can select one of the following options. • Verify selected/next batch • Verify first invalid batch
	Display the batch structure of the currently selected batch. Selecting a document shows the Document View, which provides an overview of the documents inside the batch.
	Start Learnset Manager.

Navigation toolbar buttons

The navigation toolbar enables you to easily navigate through a large number of batches. You can also configure a number of batches to appear per batch page.

Button	Description
	Go to the first batch page.
	Go to the previous batch page.
	Go to the next batch page.
	Go to the last batch page.
	Refresh

Table of batches

In the table of batches, a batch is represented by a single row. In front of each batch, a symbol displays to indicate the status. When no icon is shown, the batch state is out of workflow. You can select another batch or change the settings for the workflow.

Symbol	Status
	Batch is finished and ready for export.
	Batch requires a correction of the classification results.
	Batch requires a correction of the extraction results.
	Batch is locked and unavailable, as it is in use by another application. Therefore it cannot be opened for correction.
	Batch contains documents with exception statuses. When it is unavailable, it needs to be released before you can work on it again.

Batch view columns

The batch list can be sorted by each column. The table columns display the following information about the batch.

Batch ID

A number that can be used to uniquely identify the batch.

State

An integer between 0 and 999 that indicates the progress of batch processing. The state also indicates whether the batch is ready for verification.

Priority

An integer between 1 and 9 that indicates how urgent it is that a job be finished, where 1 is the highest priority (very urgent) and 9 is the lowest.

Name

An arbitrary name that is easier to read than the batch ID. Because the name is optional, it might be missing.

Folders

Documents in a batch can be grouped in structures called folders. The value in this column indicates the number of folders inside the batch.

Document

The value in this column indicates the number of documents inside the batch.

Client

The owner of the Intelligent Capture license. Contains N/A; not in use yet.

Last user/Module

The computer name of the operator who has processed the batch before, and the name of the application that most recently processed the batch.

Last Access

Displays the date when the batch was last processed.

External Group ID

The Group ID which has been assigned to a batch is relating to security. Batches can be assigned to user group via a unique ID.

External Batch ID

The name of the Batch Group. This can be used to represent any piece of information you would like to associate with batch. For example, external system ID and storage box ID.

Transaction ID

The Transaction ID assigned to a batch. This allows the developer to synchronize a newly created batch of documents with another external system. It can be used to identify originators of batch of documents.

Transaction Type

The Transaction Type assigned to a batch. This allows the developer to synchronize a newly created batch of documents with another external system. It can be used to identify the types of documents (such as invoices and claim forms) in batches or source of document (such as email or a scanned document).

Note: The External group ID, External Batch ID, Transaction ID, and Transaction Type columns are not displayed by default.

Sort in Batch View

You can sort any column in Batch View. To sort any item, complete the following step.

- Click on the title of the column.

  Batches sort according to their position on the list. If you select the first batch, and then click the **Batch column** label, it moves to the bottom of the list. For other items, the numbers toggle between ascending and descending order, whether numerical or alphabetical.

Select a batch in Batch View

To select a batch in the table of batches, complete the following step.

- In the table of batches, select a batch by clicking on it. You can then move through the list using the following keyboard commands.
  - To move to the first document, press `HOME`.
  - To move to the next document, press the arrow down key.
  - To move to the previous document, press the arrow up key.
  - To move to the last document, press `END`.
  - To move one page down, press `PAGE DOWN`.
  - To move one page up, press `PAGE UP`.

Leave the Batch View

To leave the Batch View and switch to another view, use one of the following keyboard commands.

- To verify the selected batch, press `CTRL+2`.
- To view the selected batch, press `CTRL+3`.

About Document View

Document view displays the batch structure of the currently selected batch. Selecting a document provides an overview of the documents inside the batch. You can use Document View to investigate the documents in a selected batch.

Open Document View

To open Document View, complete the following step.

- To select Document View, click the **Show Selected Batch**  icon.

Document View keyboard shortcuts

The following keyboard shortcuts are available in Document View.

Keyboard Shortcut	Command
`CTRL + P`	Print
`CTRL + ALT + HOME`	First Document
`CTRL + ALT + PAGE DOWN`	Next Document
`CTRL + ALT + PAGE UP`	Previous Document
`CTRL + ALT + END`	Last Document

Keyboard Shortcut	Command
CTRL + 8	Append Document
CTRL + 9	Cut Document
CTRL + ENTER	Accept/Reject next unsure page.
CTRL + SPACE	Select next unsure page.
CTRL + 1	Display the Batch View.
CTRL + 2	Start the verification of the selected batch.
CTRL + 3	Display the selected batch in Batch View.
CTRL + N	Manually restore input focus without using the mouse.
CTRL + ADD	Zoom in
CTRL + SUBTRACT	Zoom Out
CTRL + LEFT	Move Image to Left
CTRL + RIGHT	Move Image to Right
CTRL + UP	Move Image Upwards
CTRL + DOWN	Move Image Downwards
CTRL + R	Rotate
CTRL + HOME	First Page in Document
CTRL + PAGE DOWN	Previous Page in Document
CTRL + PAGE UP	Next Page in Document
CTRL + END	Last Page in Document
CTRL + M	Show selection context menu
CTRL + Z	Undo

Keyboard Shortcut	Command
F7	Option available in Advanced Verifier. Reclassify manually.
F3	Option available in Advanced Verifier. Show last verified document.
F8	Option available in Advanced Verifier. Get last value for selected field.
F9	Move to exception state.
CTRL + E	Release exception batches.
CTRL + L	Option available in Advanced Verifier. Apply local extraction.
CTRL + A	Option available in Advanced Verifier. Add document to Learnset.
CTRL + T	Correct tables.
CTRL + Q	Option available in Advanced Verifier. Switch table highlighting.

Main toolbar buttons

The toolbar provides quick access to some frequently used commands.

Button	Description
	Display a property sheet where you can configure Verifier.
	Display a property dialog where you can configure the batch filtering conditions.
	Display the Batch View.

Button	Description
	Start the verification of the currently selected batch. Depending on the batch state, the batch is either displayed in the classification window or in the indexing window. A drop-down list allows users to verify the selected batch or verify the next invalid batch.
	Display the available filters for the batch structure. You can select from among the following options. • All documents • Documents to Classify • Index Documents to Classify • Documents to Index
	Start Learnset Manager.
	Display the first page of the selected batch or a single page of the selected document.
	Display the first two pages of the selected batch horizontally or the first two pages of the selected document.
	Display the first three pages of the selected batch horizontally or the first three pages of the selected document.
	Display the first two pages of the selected batch vertically or the first two pages of the selected document.

About the batch structure area

In the batch structure, Verifier displays a hierarchical representation of the batch contents.

The levels of this hierarchy are:

- Batch
- Folder
- Document.

For each document entry, Verifier provides the following information.

- **ID**. A number that can be used to uniquely identify the batch, folder, or document.
- **State**. An integer value between 0 and 999 that indicates the progress of batch processing. The batch state is calculated from the states of its folders. It corresponds to the lowest value of all folder states. The folder state is in turn calculated from the states of the documents. It corresponds to the lowest value of all document states.
- **Name**. An arbitrary batch or folder name that is easier to read than the ID. Because the name is optional, it might be missing.
- **Document Class**. A document's classification result. This entry might be missing if the document has not been classified.

Navigate in the Document View

To navigate in the batch structure, choose from among the following keyboard commands.

- To move to the first document, press `CTRL+ALT+HOME`.
- To move to the next document, press `CTRL+ALT+PAGE DOWN`.
- To move to the previous document, press `CTRL+ALT+PAGE UP`.
- To move to the last document, press `CTRL+ALT+END`.
- To expand or collapse a folder, double-click on it, or click the plus sign or minus sign next to it.

About splitting and appending documents

In the document list, you can split multi-page documents into separate documents, with the exception of the first page of a document which cannot be split. You can also merge consecutive pages of documents into one with multiple pages.

Split a multi-page document

To split a multi-page document, complete the following steps.

1. Select **View > Show Selected Batch > All Documents**.
2. In the document list, click the multi-page document.
3. Right-click the second page, then select one of the following options.
   - Select **Cut pages into a new document**. The document is now split into two documents.
   - Select **Cut pages into new document keeping cover page**. This option splits a single document into several smaller documents and corresponding TIFFs. It also includes the cover page of the original document as the cover page for the newly created documents.

   Note: If you have changed the list sorting, such as switching the batch ID to descending order, the cut operation is not available.

   Note: This option is not available until you have marked a page as a cover page. You can do this by right-clicking on the first page of the document and selecting **Mark as cover page**.
4. Optional. If you have changed the list sorting, complete one of these options when following prompt displays:

   ```
   Would you like to switch back to the original sequence of the documents in batch?
   ```

- Click **No** to keep the sorting and disregard the cut operation.
- Click **Yes** to revert to the original sorting and cut the document.

Append two documents

To append a document to another, complete the following steps.

1. Select **View > Show Selected Batch > All Documents**.
2. Select the document to append to the previous document.
3. Right-click the document.
4. Select **Append this document to previous one**.

   Note: If you have changed the list sorting, such as switching the batch ID to descending order, the append operation is not available.

   The document now appears in the list as a multi-page document.
5. Optional. If you have changed the list sorting, complete one of these options when the following prompt displays:

   `Would you like to switch back to the original sequence of the documents in batch?`

   - Click **No** to keep the sorting and disregard the append operation.
   - Click **Yes** to revert to the original sorting and append the document.

Viewer toolbar buttons

The viewer toolbar allows you to adjust the magnification used to display documents using the following commands.

Button	Description
	Fits the document to window height.
	Fits the document to window width.
	Provides the best fit for an image.
	Zooms in.
	Zooms out.

About the document area

The document area shows the first page of the document that has been selected in the batch structure.

It is possible to default Verifier to display a specific page of each document instead of the first one. For more information, refer to the Scripting Reference Guide section on the VerifierFormLoad event.

About the verification view classification window

Verification involves taking a document that has been processed or partially processed, checking the processing results, and correcting any errors.

When you open verification view, the classification window displays automatically if the next document that is to be verified requires a correction of the classification result. Whether this is the case depends on the state of the document.

Open the verification view classification window

To display verification view, complete the following steps.

1. Select a batch from the list that requires verification.
2. Click the **Verify Selected Batch** button.

Classification window keyboard shortcuts

The following keyboard shortcuts are available in the classification window.

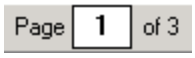
Keyboard Shortcut	Command
CTRL + ALT + HOME	First Document
CTRL + ALT + PAGE UP	Previous Document
CTRL + ALT + PAGE DOWN	Next Document
CTRL + ALT + END	Last Document
CTRL + 1	Show Batches
CTRL + 2	Verify selected batch.
CTRL + 3	Show selected batch.
CTRL + N	Restore focus
CTRL + ADD	Zoom In
CTRL + SUBTRACT	Zoom Out
CTRL + LEFT	Move Image to Left
CTRL + RIGHT	Move Image to Right
CTRL + UP	Move Image Upwards

Keyboard Shortcut	Command
CTRL + DOWN	Move Image Downwards
CTRL + R	Rotate
CTRL + HOME	First Page in Document
CTRL + PAGE DOWN	Previous Page in Document
CTRL + PAGE UP	Next Page in Document
CTRL + END	Last Page in Document
CTRL + M	Show Selection context menu.
F3	Option available in Advanced Verifier. Show last verified document.
F8	Option available in Advanced Verifier. Get last value for selected field.
CTRL + L	Option available in Advanced Verifier. Apply local extraction.
CTRL + J	Increase image area.
CTRL + K	Decrease image area.
CTRL + A	Option available in Advanced Verifier. Add document to Learnset.
F7	Reclassify document manually.
F9	Move to exception state.
CTRL + E	Release exception batches.
CTRL + T	Available in both Verifier and Advanced Verifier, but enhanced functionality only available in Advanced Verifier. Correct Table.
CTRL + Q	Option available in Advanced Verifier. Switch table highlighting.
CTRL + Z	Undo

Verification view classification window toolbar buttons

The toolbar provides quick access to some frequently used commands.

Button	Description
	Display the Batch View.
	Verify the selected batch.
	Display the Document View.
	The scope of this command depends on the Exception Mode set on the Options menu. Two options are available: Move Document to Exception State or Move Batch to Exception State. Clicking the arrow next to this button displays a list of exceptions. You can use these exceptions if you cannot correct a document at all—for example, because it belongs to none of the defined classes. Check with your administrator to determine which exceptions to use. Note that in order to avoid selection conflicts, only the toolbar button provides a list of exception handling states to choose from. The selection made here will also apply if you move a document to exception state by selecting the appropriate option within the Options menu.
	Fit the current image to the height of the window.
	Fit the current image to the width of the window.
	Fit the current image to the width or height of the window for maximum enlargement.
	Zoom in.
	Zoom out.
	Display the first document in the batch.
	Display the previous document in the batch.

Button	Description
	Display the next document in the batch.
	Display the last document in the batch.
	Rotate the current document clockwise.
	Display the first document in the batch and switches the application to Browsing Mode.
	Display the previous document in the batch and switches the application to Browsing Mode.
	Display the next document in the batch and switches the application to Browsing Mode.
	Display the last document in the batch and switches the application to Browsing Mode.
	Display the first page of the document if the current document has more than one page.
	Display the previous page of the document if the current document has more than one page.
	Enter a page number in order to navigate directly to it. All invalid entries, for example, alphabetical characters and page numbers out of range, are ignored, and the page number is reset to the currently displayed page.
	Display the next page of the document if the current document has more than one page.
	Display the last page of the document if the current document has more than one page.

About the class selection list

This box shows the classification result of the current document. If you open the list, you see all available classes.

The list entries represent the classes assigned to the current project or user and are controlled by the Verifier Classify script event.

If no result could be determined, the box shows as empty.

Set or change a classification result

To set or change a classification result, make sure that you are not in browsing mode, then complete one of the following options.

- Click on the arrow on the right side of the list box to open the list, and then select a class.
- Use the arrow keys to browse through the list of classes. The entries in the list are sorted alphabetically.
- If you know the correct class name, you can type its first characters and wait until the system automatically displays the full class name.

Select a class result using advanced classification

To select a class result using advanced classification, complete the following steps.

1. Click the **Classification Matrix** button.

   Note: This button is not available until you have checked **Enable advanced classification in Verifier** on the Verifier Mode tab in Designer.

   A list opens containing one or more classes if the result could not be determined with 100% certainty. If more than one class is in the list, the class entry is determined by probability. The class with the highest probability is at the top of the list.
2. Select a class for the current document and then click **OK**.

About the verification view indexing window

The indexing window displays fields and documents specific to your organization. The layout of the window can be customized by an application designer.

The indexing window automatically displays the next document that requires a correction of the extraction result.

Open the verification view indexing window

To display verification view, complete the following steps.

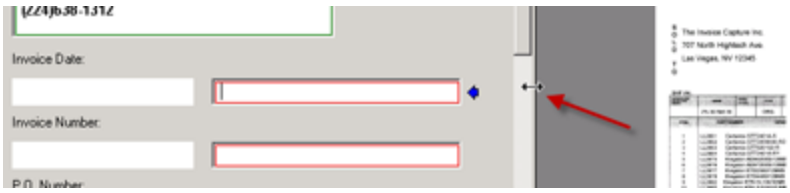
1. Select a batch from the list that requires verification.
2. Click the **Verify Selected Batch** button.

Increase or decrease the image area

To increase or decrease the image area, complete the following step.

- Drag the vertical split bar between the image area and field area either to the right or left.

  Note: This option is only available in the extraction verification view.



Indexing window keyboard shortcuts

The following keyboard shortcuts are available in the indexing window.

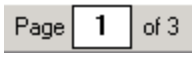
Keyboard Shortcut	Command
CTRL + P	Print
CTRL + ALT + HOME	First Document
CTRL + ALT + PAGE UP	Previous Document
CTRL + ALT + PAGE DOWN	Next Document
CTRL + ALT + END	Last Document
CTRL + 2	Verification Mode
CTRL + 1	Batch List View
CTRL + 3	Document Separation Mode
CTRL + N	Restore focus
CTRL + L	Apply local extraction
CTRL + A	Add document to Learnset
CTRL + J	Increase image area
CTRL + K	Decrease image area
F9	Move to exception state
CTRL + E	Release exception batches

Verification view indexing window toolbar buttons

The toolbar provides quick access to some frequently used commands.

Button	Description
	Displays the Batch View.

Button	Description
	Displays the Verification window.
	Displays the Document View.
	Starts Learnset Manager.
	Click the down arrow next to the button for a list of exceptions. You can use these exceptions if you cannot correct a document at all, such as when the required data is illegible. Please check with your supervisor to determine which exceptions to use.
	Marks all areas on the current document that have been used to fill the fields. If the result is valid, the area is highlighted in green. If the result is invalid, the area is highlighted in red.
	Marks only the area on the current document that was used to fill the field that is currently selected in the field area. If the extraction result is valid, the area is highlighted in green. If the extraction result is invalid, the area is highlighted in red.
	Marks the area that was used to fill the field that is currently selected in the field area. This area either appears in green or in red. In addition, all other areas that were taken into account to fill this field are highlighted in yellow.
	Fits the current image to the height of the window.
	Fits the current image to the width of the window.
	Fits the current image to the width or height of the window so that maximum enlargement is obtained.
	Zooms in.
	Zooms out.
	If this button appears pressed down, the application always displays the document area that is associated with the currently selected field.

Button	Description
	Keeps the established zoom settings on each document you view in the batch.
	Click to view the first page of the document. This button is only enabled if the current document has more than one page.
	Displays the previous page in document. This button is only enabled if the current document has more than one page.
	The text field allows you to enter a page number in order to directly navigate to that page. All invalid entries, such as alphabetic characters or page numbers out of range, will be ignored and the page number will be reset to the currently displayed page.
	Displays the last page in document. This button is only enabled if the current document has more than one page.
	Displays the next page of the document. This button is only enabled if the current document has more than one page.
	Rotates the current document 90 degrees clockwise.
	Displays the first document in the batch, and the application switches to Browsing Mode.
	Displays the previous document in the batch, and the application switches to Browsing Mode.
	Displays the next document in the batch, and the application switches to Browsing Mode.
	Displays the last document in the batch, and the application switches to Browsing Mode.
	Applies local classification and extraction to the current document.
	Adds current document to local Learnset.

Button	Description
	Starts document learning.
	Starts table correction.

Support of the mouse wheel

Verifier supports the mouse wheel when validating documents in document verification mode. Mouse wheel rolling has the following effect depending on where the mouse cursor is or where the keyboard focus is.

Case	Wheel rolling effect
Input focus is in a multi-line header field.	Scrolls between lines of the header field.
Input focus is in a single line header field or at the first line / row of any field (scrolling up only) or at the last line / row (scrolling down only).	Scrolls the entire verification form.
Input focus is in a table field.	Scrolls between table rows or between multiple lines of the currently selected table cell (when multi-line).
Mouse pointer is in the Document Viewer area.	Scrolls the currently viewed page image up and down.

Form area

A form has three main elements: a label, a viewer, and a field.

A field might be either a text field, table field, checkbox, list box, or Yes/No field. A form may also contain buttons. A form can contain any of the following elements.

Form fields

These are controls that are used to display and edit extracted data and to enter data during manual indexing. You can use form fields to create checkboxes and combo boxes.

Checkboxes

These toggle selections of data input, such as On/Off or Yes/No. Checkboxes are derived from form fields. You can set up the caption with the text desired and select the default view.

List boxes

These contain a selection list to use when verifying an item on the document. Used during manual verification, this selection works with automatic completion.

Labels

Labels are captions that help users to identify form fields, as well as viewers and tables.

Viewer

A viewer contains snippets of document areas, normally those that were extracted to fill fields or tables.

Buttons

Buttons fire actions for a new script event.

Tables

Relevant when table extraction is configured. The Verifier form supports multiple tables. However, even if you defined multiple tables, you can only display the first table on the verification form. You can display different tables on different forms.

Field area

In the field area, the following icons are used to indicate the nature of the field.

Icon	Description
	Indicates the currently selected field.
	Indicates a smart index field that can be filled by a database lookup. This icon is visible only when a smart index field is selected.
	Indicates a smart index field that can be used to start a lookup. This icon is visible regardless of whether a smart index field is selected.
	Indicates a valid extracted field.
	Indicates a field that needs to be validated because it was extracted with low confidence.

Navigate the field area

To navigate the field area, select one of the following options.

- Use the mouse. This method does not affect the validation state of a field.
- Press the `TAB` key. This method gets you to the next field, but not to the next document. This method does not affect the validation state of a field.

  Note: The order that the TAB key moves through the form is part of the form's design.
- Press the `SHIFT+TAB` key to go to a previous field.

- Press the ENTER key. This method validates the entire field or the next invalid character within a field. Once the field is corrected, it is validated and the focus moves to the next field that requires correction. This field may also be within another document.

About the verification view document area

The document area shows the currently selected document or page along with highlights.

- Red areas indicate an invalid result.
- Green areas indicate a valid result.
- Yellow areas were considered as candidates, but another candidate seemed more likely. If the extraction result is invalid or wrong, these areas may point to the correct indexing data.

Note: In practice, red, green, and yellow areas never appear in the same document.

Navigate the document area

To navigate the document area, choose one of the following options.

- To highlight the entire document table, click the square in the upper-left corner of the table field.
- To highlight a document column, click the column label of a table field.
- To highlight a document row, click the row label of a table field.
- To highlight a document cell, click the cell of a table field.

  Note: Valid areas are green and invalid are red. These areas may also contain validity icons, which are green check marks for valid fields or red Xs for invalid fields.

About tables in the document area

Only one table will display per verification form, even if you are able to define multiple tables. However, you can display different tables on different forms.

If you only need to verify certain columns in a table, you can make the other columns invisible. All invisible columns must be valid for the entire table can be valid.

For a large document with many line items, you can detect and view the location of all the extracted line items that are currently shown within a table field.

About the current input area

The current input area provides a large editing box and shows the following enlarged information for the currently select field.

- A snippet that shows an enlargement of the document area that was used to fill the field.
- The extracted data. Color coding is used in the same way as in the field area. You can edit the data here.

About the user info area

The user info area consists of three fields that display the following information (from left to right).

- The name of the currently selected field.
- If the current field is invalid, the reason is displayed. If the current field is valid, the field is normally empty.
- The classification result of the current document.

Print a document

To print a document, complete the following step.

- Select **File > Print**.

  Note: This function is available in all modes of Verifier (classification verification, extraction verification and document browsing mode) with the exception of batch browsing mode.

Printing verified data content

The amount of data on a printed form can be configured from the Print Setup dialog box. To access the Print Setup dialog box, go to File > Page Setup.

You can save your Verifier settings if you would like your current page setup preferences to be restored next time you start Verifier.

When you have configured your print settings, Verifier prints the desired field names (using display name property for each printed field) and the textual content of the fields.

The order in which the fields are printed is defined by the order of the custom fields configured in the Form Design mode of the Designer application.

In addition to the content of the fields, Verifier also prints the document file name and currently assigned document class name in the header of the printed information.

The following print setup options are available.

Option Name	Description
Print image	When selected, also prints pages of the document file (image).
Print form	Activates printing of the verification form (turned on by default). The following options below are enabled only if the present one is activated.
Print hidden fields	When selected, Verifier prints not only the fields visible on the current verification form, but all the fields available in the loaded document.

Option Name	Description
Print table fields	When deselected, Verifier does not print table fields (this option might be useful for quick printing of documents with long tables).
Print column header on each printed table page	Enabled only if the **Print table fields** option is turned on. When turned on (default), Verifier prints column header on each page (this option is useful for printing of long tables).

About working with Verifier and Advanced Verifier

This chapter provides step-by-step instructions for the main tasks that can be carried out with Verifier. In Verifier, you can perform the following three main tasks:

- Correction of page separation
- Correction of classification
- Correction of extraction

The correction of extraction results comprises the extraction of document data into form fields as well as into tables. In general, consider the following factors.

- Documents that have not been classified are displayed in the classification window. Once you have assigned a class, the indexing window displays. All fields are empty and you need to do the indexing manually.
- Documents that have been classified correctly, but which have invalid extraction results, display in the indexing window. You need to correct the extraction results.
- Documents that have been classified incorrectly are displayed in the indexing window. Press F7 to open the classification window. Correct the class and confirm by pressing ENTER. This displays the indexing window. Usually the fields will be empty because documents belonging to different classes normally do not have the same set of fields. In most cases you need to do the indexing manually.

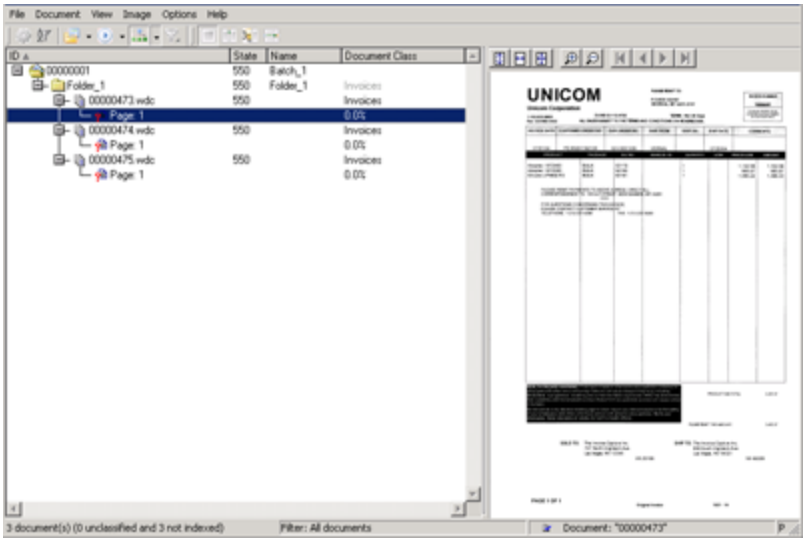
Page separation workflow in Verifer

About manual correction of automatic page separation

If, during the automatic page separation process in Runtime Server, there was at least one unsure page-level decision for a batch of documents, the whole batch receives state Failed page separation. Such a batch is supposed to be manually reviewed and, if required, corrected in Verifier.

The  icon shown indicates that this batch requires manual validation of the page separation's correctness.

You can correct the automatic page separation results in the Document Browsing mode of Verifier. When the next batch is opened, the system automatically displays the first unsure split or merged page.



The available options for automatic page separation are listed below.

- Toggle the unsure status.Select the **Accept / Reject Next Unsure Page** menu command or click CTRL+ENTER. This command sets the page to the manually accepted state or to the manually rejected state, respectively. There are three different states of page correction status: blue page icon for extracted with high confidence by the engine, blue page icon with a red question mark for extracted with low confidence by the engine (unsure) and blue page icon with green check sign for manually accepted / corrected by the Verifier user. These states are retained after the user closes the batch in Verifier and can be reviewed by other users. If all pages of a document become accepted (the pages extracted with high confidence are accepted by default), the document is redirected to successful page separation state.Refer to “240” in the previous example. Otherwise, if at least one of the document’s pages becomes manually rejected, the entire document receives the lowest page separation failed state that is configured in Verifier settings.Refer to “215” in the previous example.
- Split the document into two separate documents.Select the **Cut Document** menu command or click CTRL+9. The “top document” receives all the pages above the selected one while the “bottom document” receives all the pages below, including the selected page. In this case, the page correction status is automatically applied to the selected page and the preceding page. If you split previously merged documents,the original document names are restored.
- Merge selected document with the previous one.Select the **Append Document** menu command or click CTRL+8. In this case, the first page of the selected document and the last page of the proceeding one,are accepted for a manual page correction status.
- Go to the next unsure page.Select the **Next unsure Page** menu command or click CTRL+SPACE. This action selects the next unsure page to verify (the one with red question mark) without changing the state of any pages.

When verifying the correctness of automatic page separation in Verifier, you can now select between different convenient page view modes, for example, having two consecutive pages displayed simultaneously.

There following page view modes are available in Verifier.

- Single page view (default).
- Two pages displayed horizontally view.
- Three pages displayed horizontally view.
- Two pages displayed vertically view.

Click View > Multi-Page View to access these modes.

In multi-page view, page presentation behavesaccording to the following modes.

- In the two-page mode, the selected page is on the right (or bottom in a vertical view), and the first image shows the previous page of the current document or the last page of the previous document.
- In the three-page mode, the selected page is in the middle, the first image shows the previous page of the current document or the last page of the previous document, and the third image shows the next page of the current document or the first page of the next document.

About manual correction of classification results

Manual correction of classification results is done if the Verifier workstation is configured with the following settings.

- Classification verification is enabled.
- Extraction verification is disabled.

To determine your settings, check the Workflow tab of the Verifier Properties dialog box.

If you do this task regularly, you may want to apply the appropriate filter in the Batch View using the menu command View > Batch Filter > Batches to verify, classification only.

Manually correct classification results

To correct invalid classification results, complete the following steps.

1. In Batch View, check the state column to find a batch you can verify.

   Note: Use the arrow keys to navigate and select a batch.

2. After you select a batch, press `ENTER` to open the verification view.

   The verification view opens in Verify Mode, with the first invalid document being displayed. The cursor is already placed in the classification list box.

3. To select a class, either:
   - Click on the arrow on the right side of the list box to open the list and then select a class.
   - Use the arrow keys to browse the list of classes and make your selection. The entries in the list are sorted alphabetically.
   - If you know the correct class name, type its first characters and wait until the system automatically displays the full class name.
4. To confirm your selection, press `ENTER`.

   The application validates this document and its state increases. The next document requiring verification is displayed automatically. Proceed as described in Step 3.

5. When all documents in the batch are validated, the application prompts you to release the batch.

6. Click **Yes**, **No**, or **Details**. Clicking on **Details** reveals more options.
   - **Verify next invalid batch on the list**. Releases the current batch and opens the next batch that needs verification.
   - **Close batch and return to the batch list**. Releases the current batch and displays the Batch View where you can select the next batch.
   - **Verify this batch with the next verification form**. Change verification forms using the next verification form.

About processing documents classified to obsolete document classes

Verifier is able to correctly process (open) documents classified to non-existing (previously removed) document classes using internally saved information about the former parent class assignment.

Note that documents classified to non-existing document classes, which can often be the case in the context of supervised learning workflow, can only be processed if their former parent class still exists in the project the document is being processed with.

This feature is very useful in context of Supervised Learning workflow (using Advanced Verifier), where the vendor class is quite often deleted (or not inserted) from the global project's configuration.

About manual correction of extraction results

Manual correction of extraction results will be done if the Verifier workstation is configured with the following settings.

- Classification verification is disabled.
- Extraction verification is enabled.

Depending on the typical problems, Overwrite Mode should be enabled (to be configured within the Fields edit mode on the General Settings tab).

To determine your settings, check the Workflow tab of the Verifier Properties dialog box.

Note that if you do this task regularly, you may want to apply the appropriate filter in the Batch View using the menu command View > Batch Filter > Batches to verify extraction only.

Correct invalid results

To correct invalid results, complete the following steps.

1. In Batch View, check the state column to find a batch you can verify, then use the arrow keys to select a batch.
2. After you have selected a batch, press `ENTER` to open the verification view.

   The verification view opens in Verify Mode, and the first invalid document displays. The application places the cursor in the first invalid field; in Overwrite Mode, the field content is also selected. The user information area contains a message indicating why the field is invalid.

Form elements and field types

A form has three main elements: a label, a viewer, and a form field. From a form field, you can select a text field or table field. Using a text field or table field, you can create checkboxes or combo boxes. The field types for validation include Read Only, Auto-completion, Multi-line, Combo Box, and Checkbox. You can also add a button to a form to fire actions.

A form can include the following elements.

Form fields

Display extracted data. You can also enter and edit data during manual indexing. You can use form fields to create checkboxes and combo boxes.

Labels

Identify form fields, viewers, and tables.

Viewers

Are sections of document areas, normally those that were extracted to fill fields or tables.

Buttons

Fire actions for a new script event.

Tables

Extracted from documents.

The following is a list of field types and their description.

Read Only

When selected, information on a field is dimmed and cannot be selected or edited.

Auto-completion

Enables you to edit text in a field by typing the first two letters of a word. Auto-completion finishes the word with the best matching candidates.

Multi-line fields

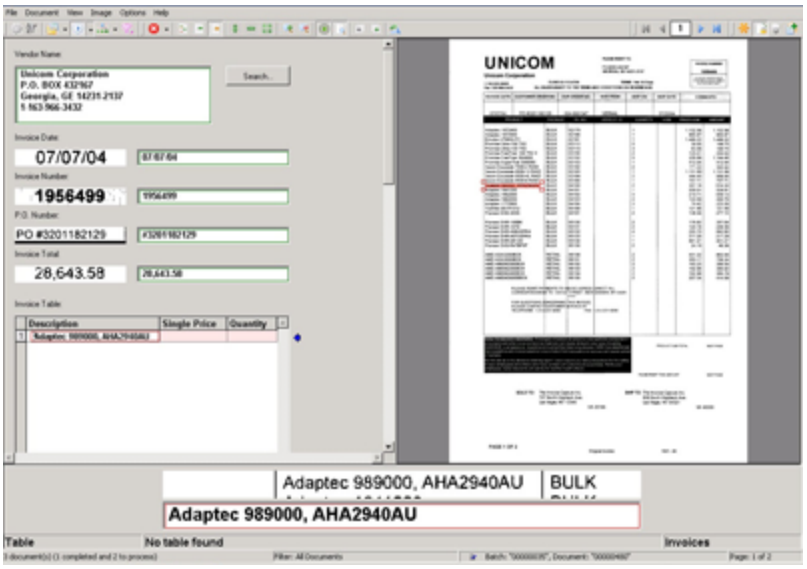
Required in the context of address analysis but can also be useful in other cases. A multi-line field enables line wrap and displays a vertical scroll bar, if required.

List box

A drop-down box that lists predefined strings related to the verification document. It can either show the nearest values automatically or show only selected values.

Checkbox

A toggle selection for one of two choices of the data input for a field (for example, Yes/No).



Note: You'll notice a Vendor Search button in many of the illustrations. This button is created and programmed in the Intelligent Capture scripting language either for the Associative Vendor search or for the Multi-column Associative search engine. In these illustrations, the button is used to quickly classify or reclassify. To learn more about the scripting language, refer to the Scripting Documentation.

About editing text fields

Verifier includes automated features for editing text fields that can speed up text entry and correction.

You can use automatic character entry, when the auto-completion is selected in the form field Properties dialog box, to edit text fields and cells. Other options for character changes include multi-line fields, combo boxes, and checkboxes. You can also insert and replace text in cells and fields, either in single words or blocks of text, using drag-and-drop or by double-clicking on the selected text.

Multi-line fields are necessary for address analysis but can also be useful in other cases. A multi-line field enables line wrap and displays a vertical scroll bar, if required. To add a new line to a multi-line field press `CTRL+ENTER`.

A Combo box lists predefined strings related to the verification document. To aid in verification, you can select from the list of strings.

The Checkbox provides an either/or option that toggles table data entry choices on and off. For example, with a Yes or No checkbox, checking Yes would bring up data entry related to the verification and unchecked for No would hide them.

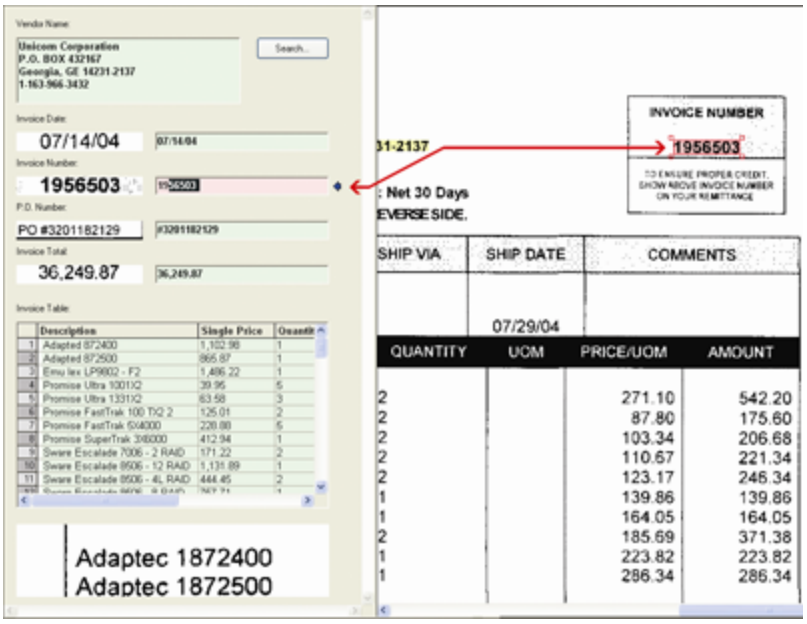
About auto-completion

Auto-completion helps to speed up typing. When you start to type, auto-text completes the word, suggesting the best match among all of the words or candidates available after OCR and Format Analysis.

For example, you can type the first two characters of a 20-character invoice. The auto-text feature finds the best matched candidate suggested by the Format Analysis engine and places it in that field. The auto-completion feature for a header field is supposed to automatically select the best candidate from the available ones, which works within Highlight Candidates mode. However, the viewer will be updated only if the candidate appears once in the document; otherwise the viewer will be blank when auto-text completes the word for the field.

The auto-selected text also appears highlighted in the original document. Select whether a single-line or a multi-line text field should be displayed. To override auto-completion, continue typing the desired text.

Note: Auto-completion does not work on formatted text and characters incorrectly read by OCR.



About inserting words in fields

To speed up verification, you can insert words to replace or append text.

The method for inserting words depends on the availability of candidates. A candidate is one that matches the learned words for that field. It will appear in green, with a border of green check marks if that visual indicator is enabled in Batch Options, when you select it after selecting the field. Non-candidates will display in orange when selected. You can insert words in fields or table cells. You can append or insert words and use the mouse to append or replace the field.

Words that are candidates for cells

If the word belongs in a cell area, you can append or replace a word in a cell. The Append feature takes the current word behind the candidate and appends the cell text. It places the text in the best location, either to the right or left of the word, and in the cell location based on the text or location of the word. The

word belonging to a cell area displays in green when selected. Or, you can replace the text.

In the search region, word candidates are all words that are not covered (by location) by other table cells and that have the same beginnings as the whole text of the cell.

Words that are not candidates for a cell

If the word does not belong to cell areas, it displays in orange when selected. Even if it is not a candidate, you can append or replace the word. Appending places the text in the best location, either to the right or left of the word, by text or location of the word. For example, a cell named “C2658” might be appended by “number.” Or you can replace the cell text and location by the text and location of a word.

Use a candidate word for a field

If a word is a candidate for a field, you can append or replace the word in a field box. A candidate is aword that matches the learned selections for the field. When you select the candidate, after selecting the field,it displays in green or issurrounded by green check marks if the visual indicator is enabled in Batch Options. The append feature appends the field text of the word left of the candidate. It places the text in the best location, either to the right or left of the word, and places the field by text or location of the word. You can also replace text. For example, a blank candidate might be replaced by “285.98.” To use a candidate word for a field, complete one of the following options.

You can insert only one candidate per field per document verification session. Furthermore, it is not possible to append words from other or remote columns. If you select words or figures from different columns, the **Append Field Text by Word**command replaces the selected field’s content. To append or replace a word in a field box, complete any of the following actions.

- To append text with the new text, click the word next to the text that you want to append. The text appears in green if it is a candidate. A box displays around the word. Double-click the box or right-click in the document and select **Append Field Text by Word**.
- To replace a word, click the desired word. A box displays around the word. Double-click the desired candidate, and drag the word to the specified field, or right-click in the document and select **Replace Field Text by Candidate** from the shortcut menu.

  Note: Make sure selected word fits the format analysis rules defined for that field. If not, the word is highlighted in orange with orange exclamation marks around it if validity icons are enabled. In this case, it would not be a good candidate for the field.

Use a word that is not a candidate for a field

Even if the word does not belong to any candidates for the field, you can append or replace a word with a new one. Appending places the text in the best location, either right or left of the word, by text or location of the word. Or you can replace the field text and location by the text and location of a word. A word that does not belong to any candidates for that field will display in orange when selected. For example, a field named “sales total” might be replaced by “invoice total.” To use a word that is not a candidate for field, complete one of the following options.

- To append text with the new text, drag a box around the desired word. Double-click on the word in the box, or right-click in the document and select **Append Field Text by Word**.
- To replace text, select the word with the mouse. A box will appear around the word. Double-click it, or select **Replace Field Text by Candidate** in the shortcut menu.

  Note: You can insert only one candidate per field per document verification session.

  Note: Make sure that this word fits the format analysis rules defined for that field. If not, the word is highlighted in orange (and with a border of orange exclamation marks if validity icons are enabled) to help distinguish it. If so, it would not be a good candidate for the field.

Insert blocks of text

Inserting large blocks of text with minimal mouse movement is helpful when you have multiple word data verification elements for fields such as address information or cell descriptions. Before you can insert blocks of text, first select the settings in the Workflow dialog box to immediately copy information. To insert large blocks of text, complete the following steps.

1. Click and drag over the desired text in the image viewer.
2. Release the mouse button. A rectangle appears around the text.
3. Adjust the rectangle, if necessary, by selecting the nodes at any corner.
4. Drag and drop the rectangle to the field or table cell. A copy of the rectangle appears over the field or table cell. Optionally, you can double-click on the rectangle. The text in the rectangle replaces the text in the field or table cell.

   Note: You can move or resize this rectangle by clicking in the area in the image viewer. When the rectangle appears, select the nodes to resize it, or drag it using the drag and drop method described above.

Finish the validation

1. After you correct a field, press `ENTER` to validate it.

   During validation, the field’s background color appears in yellow, and the cursor becomes an hourglass. Once the validation is finished, the cursor moves automatically to the next invalid field regardless of whether this field is still in the same or already in the next invalid document. If you leave a document this way, it is validated automatically. In the next field, proceed as described above. When all documents in the batch are validated, the application prompts you to select what to do next.

2. Click **Yes** or **No**, or click **Details**. Clicking **Details** reveals the following options.
   - **Verify next invalid batch on the list**. This releases the current batch and opens the next batch that needs verification.
   - **Close batch and return to the batch list**. This releases the current batch and displays the Batch View where you can select the next batch.
   - **Verify this batch with the next verification form**. Change verification forms using the next verification form.

About manual correction of classification and extraction results

Simultaneous correction of classification and extraction results is available if your workstation is configured with the following settings.

- Classification verification is enabled.
- Extraction verification is enabled.
- Automatic extraction after classification is disabled.

Note: If this option is enabled, extraction will be carried out automatically by Runtime Server. In this case, classification verification and extraction verification are two separate steps.

Note: To determine your settings, check the workflow tab of the Verifier Properties dialog box.

Note: If you do this task regularly, you may want to apply the appropriate filter in the Batch View using the menu command View > Batch Filter > Batches to verify.

About smart indexing

Organizations, in particular commercial ones, usually collect legions of information about themselves and everybody they do business with. Much of this information is stored in databases. There will rarely be a company without a customer database that contains addresses, contacts, and so on. Financial transactions are also recorded in databases. Databases can be an excellent support for indexing because they store related information that can easily be retrieved. During indexing, if you have one piece of information from a document, you can obtain related pieces from the database and fill the associated fields automatically. This method is called smart indexing.

Normally, smart indexing is combined with manual indexing. Some fields of a form have to be filled in manually; some fields can be filled automatically.

For example, let's assume that your organization saves information related to orders in the database of its ERP system. Every order is characterized with a unique identifier and some attributes about the supplier and the items that have been ordered. Soon after an order is placed, the ordered items are delivered, and a delivery note is attached. The corresponding invoice follows soon. Delivery note and invoice refer to the original order. They have the order's unique identifier printed on them. With this identifier, you can look up supplier information from the database when you verify the delivery note and invoice. However, new information such as the invoice date has not yet been entered into the database. This information can be supplied manually.

Use smart indexing

To use smart indexing, complete the following steps.

1. Smart index fields can be recognized by the Key icon that is displayed next to them. Select a smart index field. The field itself and all the fields that can be filled through the database lookup are marked with a yellow cylinder.
2. If the field is still empty, enter the field value. Alternatively, enter a wildcard expression, using a * to represent a sequence character or a ? to represent a single character.
3. Complete one of the following options to start the lookup.
   - If your application is configured accordingly and the field content is correct, validate the smart index field by pressing `ENTER`.
   - Press `ALT+F12`.
4. The system may respond with one of the following options.
   - If the lookup yields no results, a corresponding message is displayed. Fill the lookup fields manually. If you cannot complete the fields, send the document to exception handling.
   - If the lookup yields one result, the lookup fields are filled.
   - If the lookup yields multiple results, and this is allowed in your application, the lookup fields are filled.
   - If the lookup yields multiple results, and this is not allowed in your application, a dialog box is displayed where you can select the correct record. The lookup fields are then filled accordingly.

Check entire batches

To browse through all documents in a batch, complete the following steps.

1. In Batch View, use the status value to determine a batch you can browse through. Use the arrow keys to select a batch.
2. After you have selected a batch, press `ENTER` to open the verification view.

   The Verification View displays the first document requiring correction.
3. To display the first document in the batch, press `CTRL+ALT+HOME`.
4. You may encounter a document that has been classified incorrectly. To correct this result, press `F7` to open the classification window. To correct the class, select the corresponding entry from the list box at the bottom, then confirm by pressing `ENTER`.

   This displays the indexing window again.
5. To correct extraction results, type your corrections into the corresponding field. If a field has been changed, its state is set to invalid. Press `ENTER` to validate the field you modified, and then press `F3` to return to the document.
6. To get to the next document, press `CTRL+ALT+PAGE DOWN`. Proceed like this until you reach the last document.

About working with tables

In Intelligent Capture, it is possible that any given table was trained by the new extraction engine, Brainware Table Extraction, or the traditional Table Analysis Engine. As a Verifier user you will not know which engine was being used and you will not be able to see a difference in Verifier. The process of table fields is similar regardless of whether you are using Verifier or Advanced Verifier.

The difference between the two workflows, Verifier and Advanced Verifier, lies in the fact that tables extracted through Brainware Table Extraction can be learned and corrected differently by Supervised LearningVerifier users. Even if you're a Supervised Learning Verifier, you will not be able to train and correct traditionally created tables the same way you can a Brainware Table Extraction table.

Note that you can correct invalid cells the same way you would correct an invalid text field.

About automatic training and extraction of verified table data

Brainware Table Extraction supports automatic learning of verified table data. Brainware Table Extraction trains documents using only the information in verified table data, the content and position of every data cell.

Brainware Table extraction supports Generic Table extraction. When using Generic Table Extraction, table data is easily mapped with numerical correlation and label matching. No extra learning is needed.

Column and line extraction is not supported with the Generic Table Extraction. You must extract and correct fields manually while working with Generic Table Extraction.

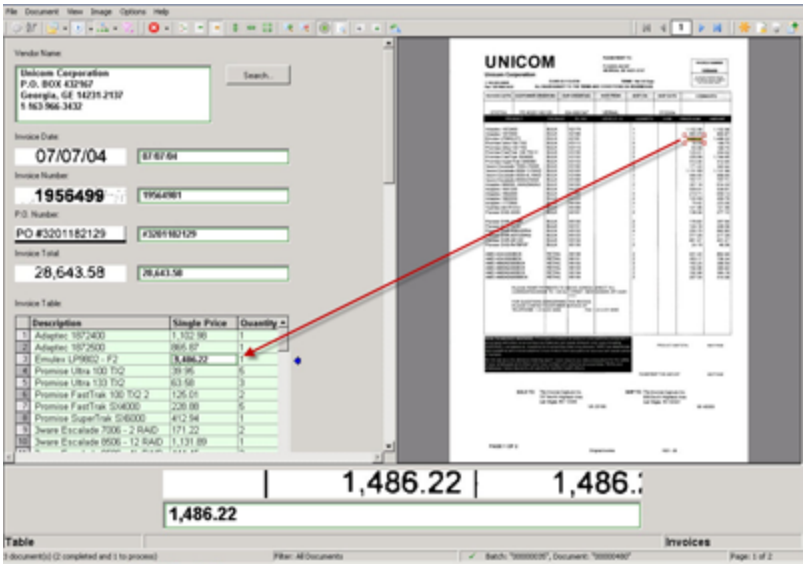
About manual training and correction methods

In the case that automatic table extraction fails to recognize the line items properly, Verifier provides several ways for convenient manual table correction.

About auto-completion with tables

Auto-completion works in table cells and with text fields. When you type two or more characters, auto-complete suggests a word or phrase for that cell.

The candidate appears in green if the field is valid and red if the field is invalid. If the visual validity icons are enabled in Batch Options, valid fields also have a border of green check marks and invalid fields have a border of red question marks. This function only works with Highlight Candidates mode.



Correcting table structure

You may also need to correct the table structure. Table rows, cells, and columns have shortcut menus with options for modifying the table structure. To invoke them, right-click on the row, cell, or column label.

The available commands are summarized in the table below.

Shortcut menu	Command	Description
Column	Unmap	Clears all data for the selected verification column and turns the state of the corresponding column of the recognized table back to "unmapped." To view an unmapped column, double-click on the table header in the verification form. All unmapped columns are highlighted in red.
Column	Map	Adds the column selected from the shortcut menu, or you can right-click on an unmapped column to map it to a column in the verification form.
Column	Swap	Exchanges the position of the current column and the one selected from the drop-down menu.
Row	Insert	Inserts an empty row above the current one.
Row	Delete	Deletes the current row.
Row	Duplicate	Duplicates the current row.
Row	Append	Appends an empty row at the bottom of the table.
Row	Merge	Merges cells in a row.
Cell	Insert Cell	Inserts an empty cell above the selected cell while shifting the cells below down.
Cell	Delete Cell	Deletes the selected cell. The cells below are shifted up.

About the rubber-banding feature

The rubber-banding feature allows you to select a block of data on a document and to place this block of values at a particular point within the table.

Note that the Table Correction mode must be switched off for the menus described in the following sections, though a mixed usage of manual and engine driven table correction is possible.

Auto-scroll with the rubber-banding feature

To auto-scroll when using the rubber-banding feature, use one of the following options.

- If the target document area is displayed only partially, such as due to the zoom level, move the mouse outside the document area while rubber-banding to scroll the document and to select the entire desired data.
- If you want to re-size a rubber-banded area, drag the corner of the rubber-band rectangle. The window will auto-scroll if you want to select values outside of the visible area.

Add column data to a table

There are two ways to add column data to a table.

Insert Column Table Data

If whole rows are missing after extraction, use this option for better accuracy. Note that the application observes the relationship between the columns and maps the values appropriately. Already extracted cell values would be shifted up or down by the insertion creating new rows. See the insertion behavior described below.

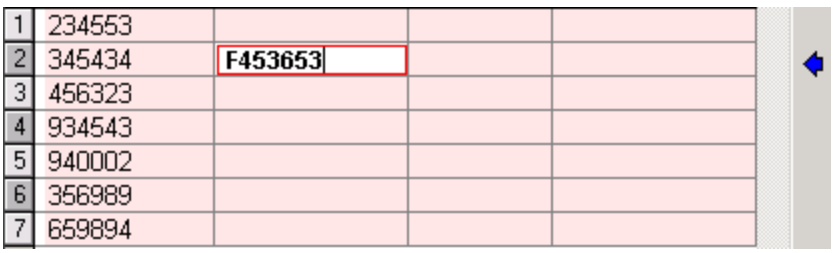
Replace Column Table Data

This is very comfortable when at least one column has been extracted by 100%, and other table columns contain random values. With this option, the column data can be added blockwise, overwriting the previously extracted entries.

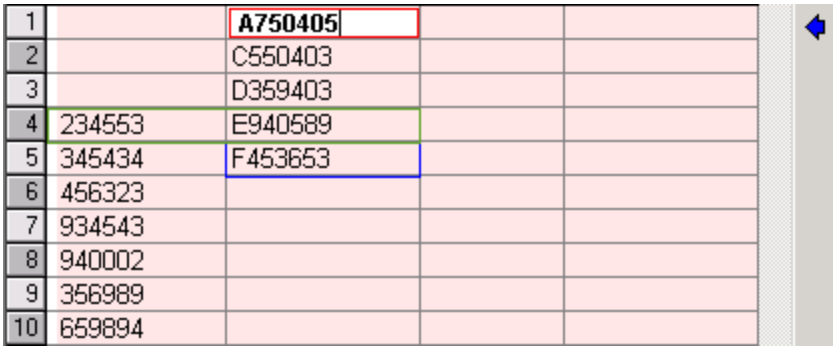
Insert column data

To insert column data, complete one of the following options.

- To insert column data above a correctly extracted and filled cell, click into the filled cell and select **Insert Column Table Data** from the menu. This creates additional rows and shifts already existing rows up. At the same time, the values will be automatically assigned to already extracted values of other columns if available (green frame in the screenshot below).



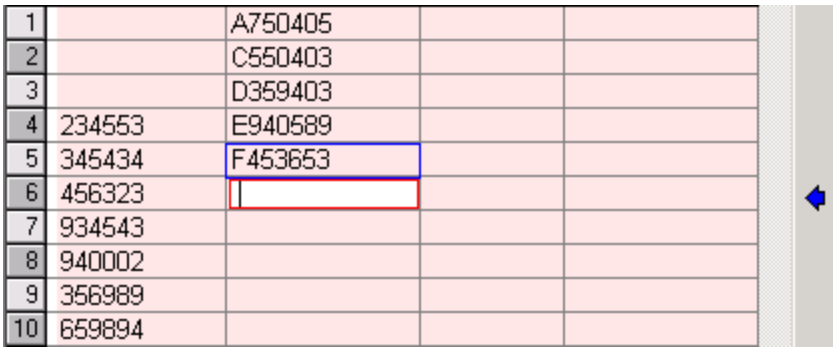
- To insert column data below a correctly extracted and filled cell, click into the next empty cell below and select one of the two options depending on whether or not you want to keep already extracted cell entries.



Replace column data

To replace column data, complete the following step.

- To overwrite already existing table cell values, click into the desired starting point cell and select the **Replace Column Table Data** option.

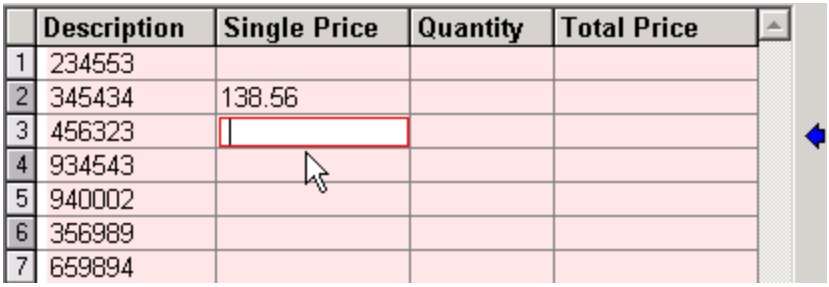


Provided that the rubber-banded area spans more lines than already contained in the table, the additional subsequent lines will be added as additional rows continuing the table.

Use the rubber-banding feature

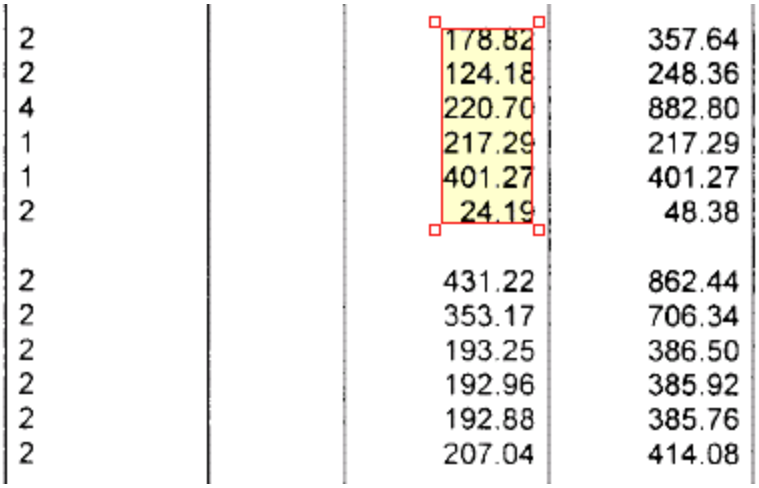
To use the rubber-banding feature, complete the following steps.

1. Place the cursor into the destination cell within the table.



	Description	Single Price	Quantity	Total Price
1	234553			
2	345434	138.56		
3	456323			
4	934543			
5	940002			
6	356989			
7	659894			

2. Draw a rubber band rectangle around the column data within the Document Viewer.



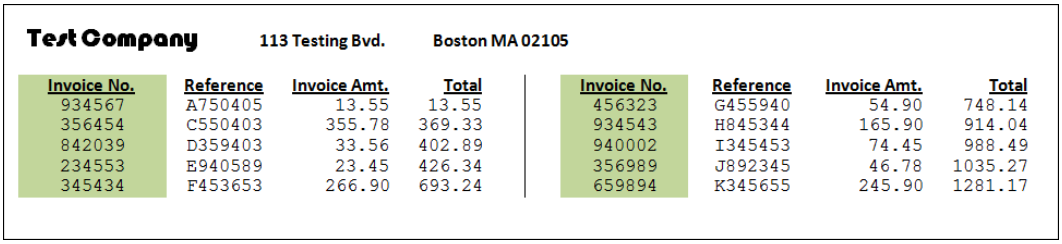
2		178.82	357.64
2		124.18	248.36
4		220.70	882.80
1		217.29	217.29
1		401.27	401.27
2		24.19	48.38
2		431.22	862.44
2		353.17	706.34
2		193.25	386.50
2		192.96	385.92
2		192.88	385.76
2		207.04	414.08

3. Right-click the selection and click **Insert Column Table Data** within the context menu or **Replace Column Table Data** depending on your task.

Use the rubber-banding feature: use case one

A use case for the rubber-banding feature is if columns of one data type are split up and placed side-by-side, or stacked, on a page as shown in the example below. To correct such a table, complete the following steps.

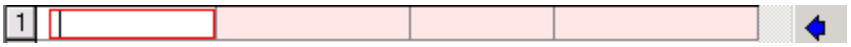
1. Place the cursor into the first line of the column.



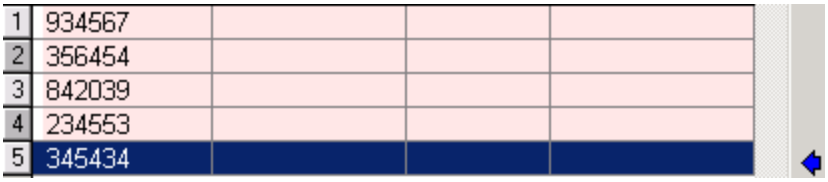
Test Company 113 Testing Bvd. Boston MA 02105

Invoice No.	Reference	Invoice Amt.	Total
934567	A750405	13.55	13.55
356454	C550403	355.78	369.33
842039	D359403	33.56	402.89
234553	E940589	23.45	426.34
345434	F453653	266.90	693.24

Invoice No.	Reference	Invoice Amt.	Total
456323	G455940	54.90	748.14
934543	H845344	165.90	914.04
940002	I345453	74.45	988.49
356989	J892345	46.78	1035.27
659894	K345655	245.90	1281.17



2. Draw a rubber band rectangle around the first block of column data on the document.
3. Right-click the selection and click **Insert Column Table Data**. Based on the document example above, the mapping result would look like the following example.



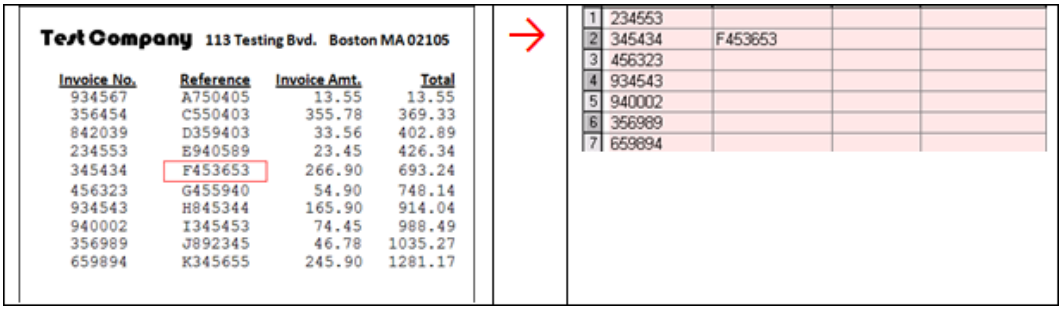
4. To append further column data, right-click the row's node of the last table row and select **Append Row**.
5. Place the cursor into the next empty cell of the desired table column.
6. Draw a rubber band rectangle around the next data block of the same column type and select **Insert Column Table Data**.
7. Proceed the same way with other columns to fill the table.

Use the rubber-banding feature: use case two

A use case for the rubber-banding feature is if column items have been extracted only partially. To correct such a table, complete the following steps.

1. Place the cursor into the already extracted cell to mark the starting point for the insertion.
2. Draw a rubber band rectangle around the column data on the document.
3. Right-click the selection and select **Insert Column Table Data**.

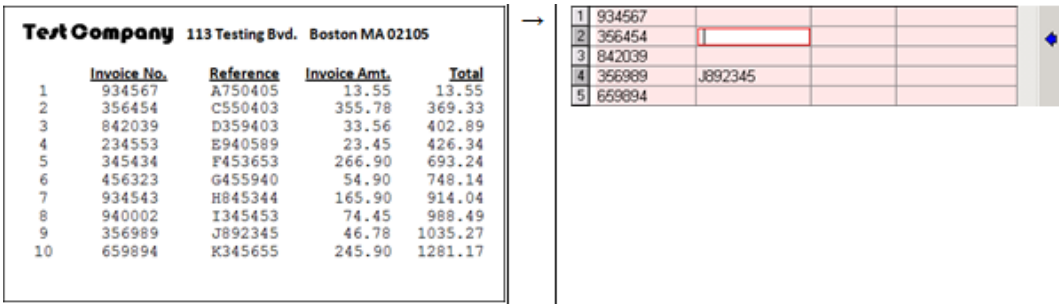
4. Now continue to map the other column data. To add the column data under the cell “F453653”, select the very next empty cell below as starting point for insertion and repeat the steps from above.



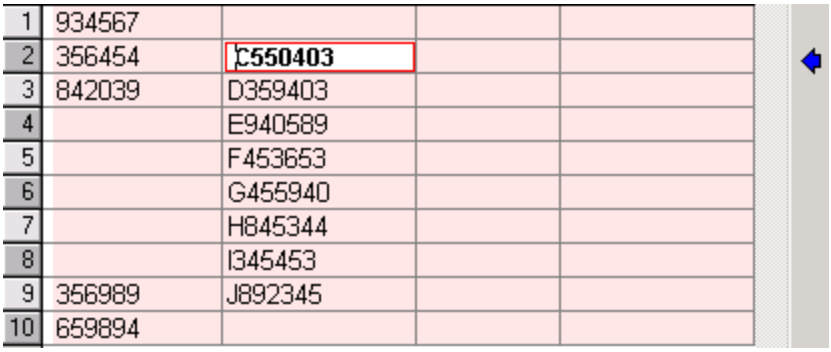
Use the rubber-banding feature: use case three

A use case for the rubber-banding feature is if there are missing column items in the neighbor column. Taken the document example above, the column data has been extracted as in the following example. From the ‘Reference’ column, only the value from line 9 is extracted, and the ‘Invoice No.’ values from lines 4 – 8 are missing. To insert the ‘Reference’ values from lines 2 – 8 as a block, complete the following steps.

1. Place your cursor into the ‘Reference’ cell in line 2.



2. Perform the steps as described in the previous use case.

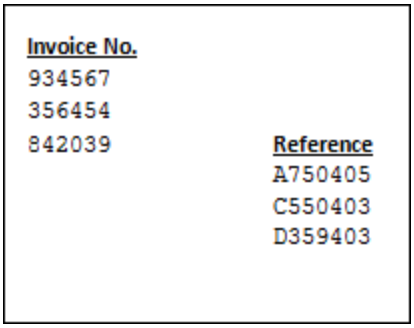


Result ‘Reference’ values where the according ‘Invoice No’ values are missing, are inserted as rows while shifting the subsequent, already extracted cell “J892345” down. Now, the missing data from the ‘Invoice No.’column can be added.

Use the rubber-banding feature: use case four

A use case for the rubber-banding feature is if you have documents where the data columns appear misaligned. To correct this situation, complete the following step.

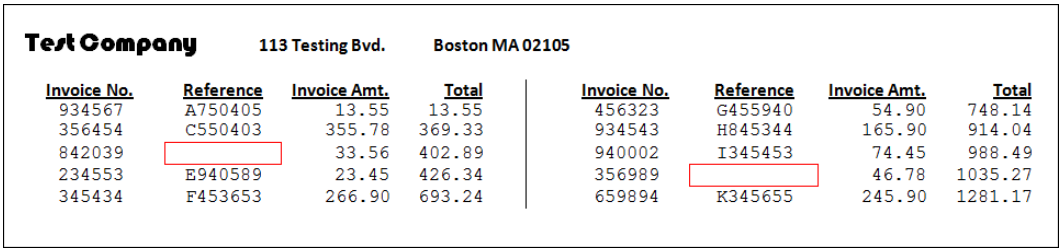
- Map the column data blockwise as described for previous use cases.



About correcting single cells

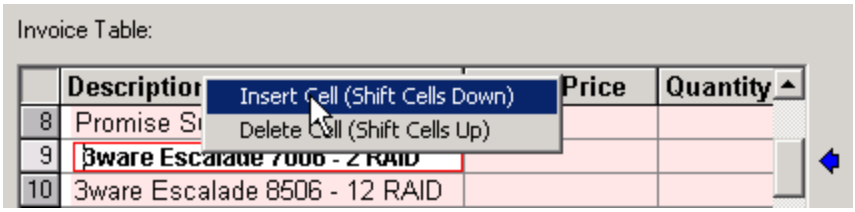
You may also need to correct the table structure if for instance an unnecessary cell has been mapped to the table or if a missing cell has to be added.

As in the example below, you may have documents where one of the line items is missing. During extraction, the values from below might be shifted up to fill the empty space.



For this, you have the possibility to add or remove single cells.

For each of the table cells, the following context menu is available.



Delete an unnecessary cell

To delete a cell, complete the following steps.

1. Click the cell within the table and place the cursor in it.
2. Right-click and select **Delete Cell (Shift Cells Up)**.

Insert a cell

To insert a cell, complete the following steps.

1. Click the cell within the table to be subsequent to the cell candidate and place the cursor in it.
2. Right-click and select **Insert Cell (Shift Cells Down)**.

Result This creates an empty cell within the table above the selected cell. Now, you can copy the desired value from the document into the newly created cell.

About table extraction and correction

The learning process for the Brainware Table Extraction engine consists of two phases.

- Learning lines
- Learning mappings of columns

These are discussed in detail in the following sections.

Note that functionality is available for the Supervised Learning Verifiers. With the Generic Table Extraction, no extra learning is needed.

Learning lines

The Brainware Table Extraction engine considers the following main types of the lines.

Primary line

A line that defines table structure. The engine applies advanced and precise similarity analysis for all primary lines. It is important that all primary lines are well-structured and that they look similar in many of the rows to extract. The engine easily supports an unlimited number of types of primary lines for one table definition. The primary line must contain at least four words. Otherwise, the engine will not learn it. In addition, the primary line must be the first line in the table row.

Secondary line

A line between primary lines. The engine applies smooth similarity analysis for these types of lines, which is possible because Brainware Table Extraction only searches the area between two neighboring primary lines. This allows the engine to extract data that varies widely, which often happens with multi-line descriptions. There is also no limit to the number of words in secondary lines, and no limit to the number of secondary lines. However, a document's page must have at least one primary line; otherwise, secondary lines on this page are not extracted.

Wrong line

A primary line that is learned as a negative line sample. In other words, all lines classified by the engine as members of one particular “wrong” line class are not extracted. In principle, it is possible to learn an unlimited number of wrong lines, though the current restriction is that this will only take effect during in-document learning. Cross-document learning (that is, learning the whole document after all the fields are completely valid) may not automatically train the wrong lines.

After it learns any type of line, the Brainware Table Extraction engine automatically creates and manages a new line class (cluster). Afterward, all lines in the document considered by the engine to be members of the line class (similar to the learned line sample) will be extracted, or not extracted in the case of “wrong” lines.

It is possible to learn an unlimited number of different line classes. However, the overall quality may suffer if too many lines are learned.

Learning lines can be applied in lines learning (or lines highlighting) mode. Mapping of the column data in the lines can be done in column mapping learning (or columns highlighting) mode. The user can switch between learning (highlighting) modes with the Switch Table Highlighting menu option in Verifier Options menu or with the context menu options Show Lines and Show Columns.

About learning mappings of columns

When learning the mapping for columns, the user trains the engine on how the data from the extracted lines must be mapped to the user's table data.

For primary lines, this mapping can be defined differently for different line classes. For example, if a user learned two different line samples that went to two different lines classes internally in one document, the user can then map “Unit Price” in the document to the “Unit Price” data column and the “Total Price” to the “Total Price” for the first line sample. For all lines of the second line type, the user can map “Unit Price” to “Total Price” and “Total Price” to “Unit Price.” For the next document, the Brainware Table Extraction engine will always use mapping rules #1 for the lines classified to the first line type and mapping rules #2 for the lines classified as the second line type.

If you have several Brainware Table Extraction tables in one class, the Learnset is shared between these tables. In other words, if you used interactive learning for one Brainware Table Extraction table, cross-document learning (which happens if the system added the document to the Learnset after document validation) is applied for all Brainware Table Extraction tables in the document.

About correcting fields in tables created with Brainware Table Extraction

Any time you train a table interactively, complete the required training first and then manually verify the table.

Brainware Table Extraction can train line types and column mapping for each type of line.

When working with interactive table extraction, learn the lines before you map the columns.

Because of the way interactive table verification works, you cannot manually delete data from a cell. Rather, if you want to discard cell data, unmap the column and re-extract the table to remap the column. Although it will seem as if you deleted the data, the data is still be there until you unmap the column.

Use the standard method for table extraction

This section describes the simplest way to use interactive Brainware Table Extraction learning. If this method does not work, proceed to the advanced method described in the following sections. To use the standard method, complete the following procedures.

1. Show the first row sample.
2. Learn mapping in the learned row.
3. Learn missing lines.
4. Learn and adjust the mapping of missing or wrong columns.

5. Manually correct the table date and validate the table.
6. Learn the document.

Step 1: Show the first row sample

1. Select your Brainware Table Extraction table by clicking any table field inside the table grid.
2. Click the **Correct Tables** button.
3. In the lines highlighting mode, use the **Learn As Row** function to show the row sample. This function automatically learns the first line as a primary line and the rest of the lines as secondary lines. This function is also available by double-clicking on the selected row area. Select the whole first row and learn it.

   Note: The visual indicators for valid, invalid and questionable table lines are the same as for header fields: valid lines have a green check mark; invalid lines have a red X, and questionable fields have an orange question mark.

4. Optional. To learn a new line as a primary line, complete the following substeps.
   1. Right-click on any line marked in gray in the Image Viewer.
   2. On the shortcut menu, select **Learn Line**.

      The learned lines change from gray to green, or to blue if the line is extracted with low confidence.
5. Optional. To learn a block of lines as primary lines, complete the following substeps.
   1. In the Document Viewer, draw a rectangular selection over the primary lines.
   2. Right-click on the selection.
   3. On the shortcut menu, select **Learn as Primary Line(s).**

      All correctly selected primary lines will be learned and highlighted in green (or blue if the line is selected with low confidence), and all other lines will be similarly extracted and displayed.

   Note: If some lines were not extracted (these lines will not be color-coded) try relearning the lines singly or in a block.

6. Optional. To learn a lines block as a table row, complete the following substeps.
   1. In the Document Viewer, draw a rectangular selection over the required multi-line (or single-line) table row.
   2. Double-click or right-click on the selection.
   3. From the shortcut menu, select **Learn as Row**.

      Note: If some lines were not extracted, repeat the procedure described above.

      Note: Do not try to learn the rest of missing secondary or primary lines now. This is because mapping is defined on the basis of line type. If you would train all different line samples now, you would need to learn the columns mapping separately for every line class. In order to reduce time to train the table, first learn the column mapping for the row you just learned. If you then want to learn another line sample, the engine will apply the existing mapping rules for the newly learned row automatically.

      Note: Green highlighting indicates a line is extracted with high-confidence; blue highlighting - with low-confidence. If the confidence for a blue line is less than 0.3 (moving the mouse cursor over the highlighted lines shows the confidence value as a tool-tip) then the lines will not be extracted. Blue highlighting has also the following important meaning: this line can be trained by the engine as a new line class.

All correctly selected primary lines will be learned, and all other lines are similarly extracted and displayed.

Step 2: Learn mapping in the row you learned

1. Switch to the columns highlighting mode now (using `CTRL+Q`) and mark the location of your first cell item in the row you learned.

   The system displays a special mapping control asking for the data column to extract the data to.
2. Select the required data column by double-clicking on it.
3. Repeat this step for the rest of the cell items in the first row.

Step 3: Learn missing lines

1. Switch back to the lines highlighting mode.

   Note: Pressing `CTRL+Q` switches the highlighting between three modes: cells, lines, and columns. Press `CTRL+Q` twice to switch from columns learning to lines learning.
2. Mark the next missing row and learn it as described before.
3. Repeat this step for all rows on all pages where something is missing. Go to the next step only after you are sure nothing is missing.

Step 4: Learn and adjust the mapping of missing or wrong columns

1. Return to columns mapping learning mode and look for wrong or missing mapping. Correct any missing mapping.
2. If you can't map the missing columns, switch back to the lines highlighting mode and try to learn the row where the mapping is missing.
3. Switch to columns highlighting. If the mapping is still missing, mark the missing part and map it.

   Note: The Brainware Table Extraction engine may determine the mapping automatically.
4. Repeat these steps until the data is completely extracted or cannot be learned correctly. (There is always a chance that you will not get 100 percent extraction results.)

Step 5: Manually correct the table data and validate the table

- Now switch to cells highlighting mode and manually correct missing data, OCR errors, and so on.

  Note: Do not use interactive learning anymore because every Brainware Table Extraction learning action will reactivate extraction and will replace all your manual input at one time.

Step 6: Learn the document if required

After table learning and validation are complete and the rest of the document's fields are validated, you may want to add this document to the Learnset and then learn it (this is "cross-document" learning, in contrast with "in-document" interactive Brainware Table Extraction learning.)

- If the system did not suggest learning the document automatically (the Add Current Document to Local Learnset toolbar button is not pushed), but you still would like to learn your table, activate learning by clicking the **Add Current Document to Local Learnset** button.

  Note: The only requirement for cross-document learning is correctness and completeness of the table data to train. This means that location and content of every cell item should be correct. Also, ideally, the content of cell items should not be formatted.

Advanced learning with Brainware Table Extraction

This section discusses the special cases in which it is necessary to use secondary lines explicitly. There are two such cases.

Case 1: Table row begins on one page and ends on the next

If a table row begins on one page and ends on the next page, you must use the Learn as Secondary Lines function (in lines learning mode) to train missing secondary lines (on the next page.) In this case, these secondary lines are placed right before the first primary line on the page. Mark all the secondary lines as before: right-click and select Learn as Secondary Lines.

Never use the Learn as Row function in this case, as this tells the engine that the first secondary line is actually a new sample of primary line. As a result, the engine may split extracted table data into new rows.

Case 2: Learning of unmapped secondary lines leads to unwanted extraction

Your project may require that data from secondary lines not be extracted. Usually, this will not be a problem, but sometimes the engine extracts the data from these lines anyway. In this case, not learning these secondary lines will prevent unwanted extractions. Use the Learn as Secondary Lines function instead of Learn as Row if you would like to learn just selected lines and not all lines that belong to the row. You can also Unlearn Line to correct or adjust the extraction.

Learn a block of secondary lines

To learn a block of secondary lines, complete the following steps.

1. In the Document Viewer, draw a rectangular selection over the required secondary lines of a desired multi-line row.
2. Right-click on the selection.
3. On the shortcut menu, select **Learn as Secondary Line(s)**.

   All correctly selected primary lines will be learned and highlighted in green (or blue if the line is selected with low confidence) and all other lines are similarly extracted and displayed. If some lines were not extracted (these lines will not be color-coded), repeat the procedure described immediately above.

Advanced learning: additional functions

About the Unmap Column method

The Unmap Column method can undo mapping for the specified cell item.

This will undo mapping for all cell items that were extracted from the lines that belong to the same line type as the cell item used to invoke the Unmap Column method.

Undo column mapping

To undo incorrect column mapping, complete the following steps.

1. Right-click on any unassigned column – they are marked in blue – or draw a rectangular selection over the cell items to be mapped to a table column.
2. On the shortcut menu, select **Undo Mapping**.

Result The previously assigned column (highlighted in red) is now unassigned (and displayed in blue.) The values are no longer extracted or in the table grid.

Unlearn a line

The Unlearn Line function can be used to discard previously applied learning for a particular line. To do this, Brainware Table Extraction uses a line sample, searches for the line type, and removes the line type from the Learnset. To unlearn a line, complete the following steps.

1. Switch to **Lines Learning** mode and right-click on the line you want to unlearn.
2. On the shortcut menu, click **Unlearn Line**.

   Unlearned lines change from green to gray.

Learn a line as a wrong line

Learning a wrong line means to train the table such that a particular line will not be extracted. This applies to other lines of the same type in the table. To learn a line as a wrong line, complete the following steps.

1. Right-click on any learned line or draw a rectangular selection over the required lines.
2. On the shortcut menu, select **Learn as Wrong Line**.

   The selected lines and similar lines to it are now highlighted in gray. Information from these lines will not be extracted.

Working with Learnset Manager

About Supervised Learning

The basic purpose of Learnset Manager is to use Supervised Learning (interactive training) to improve the quality and usefulness of your enterprise's Learnsets.

With Supervised Learning, Supervised Learning Verifiers, and Learnset Managers, you can customize your project's Learnsets by adding or subtracting documents, reclassifying them or creating altogether new classes or Learnsets and migrating documents there. They can also promote Local Learnsets to a Global Learnset so that it can be shared across the enterprise.

In general, Learnset Manager consists of:

- Creating new classes based upon documents themselves and supplier information.
- Learning documents and adding them to the Local Learnset.
- Using the Local Learnset to improve the extraction of low-quality documents.
- Maintaining Local Learnsets.
- Updating and enhancing the Global Learnset with information from the Local Learnsets.

All this is done through a simple interface that closely resembles that of Verifier and Advanced Verifier.

Although Supervised Learning was created for use with vendors' invoices, it can also be used with other types of knowledge. For example, a library might create classes based on type of material, subject matter, or author. Most of the illustrations and examples in this chapter use invoices.

Learnset Manager is a separately licensed add-in that can only be launched from Verifier. You have access to Learnset Manager mode only if you have been assigned to a group that has permission to work with the mode. There is no limit on the number of users who can simultaneously access Learnset Manager.

Start Learnset Manager

To start Learnset Manager, complete the following step.

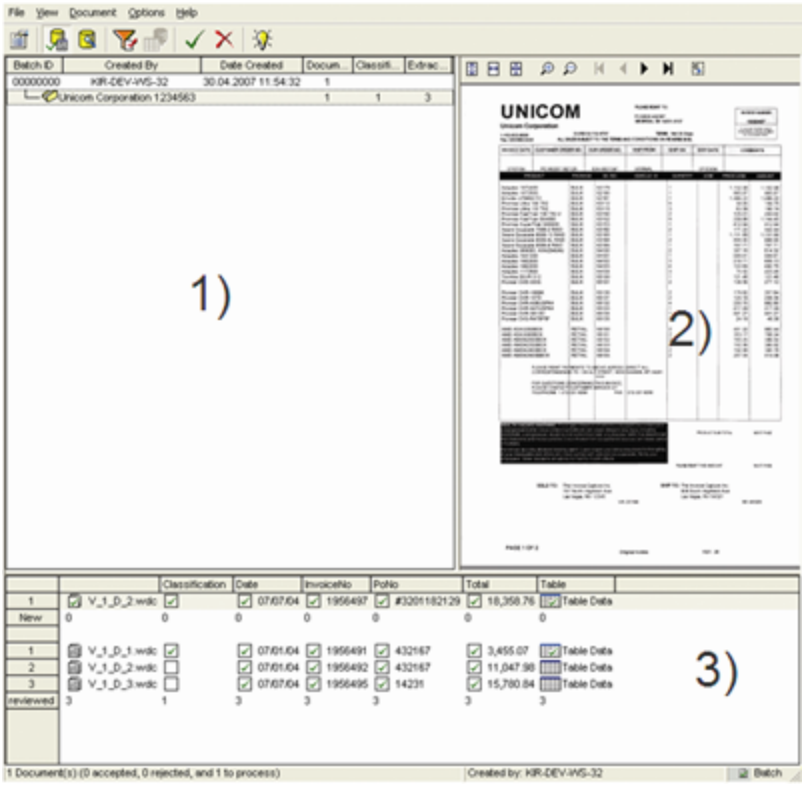
- Click the Learnset Manager button in Verifier.

Getting familiar with the Learnset Manager user interface

Learnset Manager has two basic modes, or views. These are the Accumulated Documents View (where you work with Local Learnsets) and the Global Learnset View (where you work with Common Learnsets and Global Learnsets.)

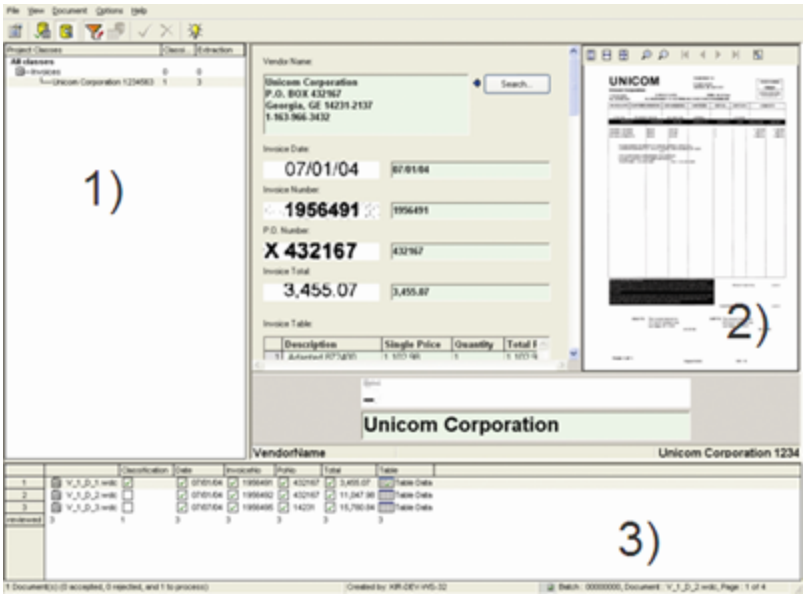
When you are working with Local Learnsets, you can further refine the appearance of the Accumulated Documents Browser when you verify documents or manually reclassify them.

Accumulated Documents browsing



1. **The Batch viewer**. Enables you to see each class in the batch you are working on. The Batch Viewer shows each class as part of the batch, the user who created the batch, the date it was created, the number of documents in the batch and in each class, the number of documents successfully classified, and the number of documents successfully extracted. You'll need to enlarge the window to see all of these categories.
2. **The Document Viewer**. As with the Document Viewer in Verifier, this window enables you to see (and therefore verify) each document in the batch you're working on.
3. **The Learning Statistics Window**. Shows the documents that have been processed by Intelligent Capture. Documents that are awaiting processing have a question mark beside them. Successfully processed documents have a check mark, while documents that failed processing have an X.

Global Learnset browsing



1. **The Batch viewer**. Enables you to see each class in the batch you are working on. Shows the classes in the Global Learnset, and the number of documents classified or extracted in each.
2. **The Document Viewer**. As with the Document Viewer in Verifier, this window enables you to see (and therefore verify) each document in the batch you're working on.
3. **The Learning Statistics Pane**. Shows the documents that have been processed by Intelligent Capture. Documents that are awaiting processing have a question mark beside them. Successfully processed documents have a check mark, while documents that failed processing have an X.

Learnset Manager keyboard shortcuts

The following menu commands and keyboard shortcuts are available in Learnset Manager.

Keyboard Shortcut	Description
CTRL + E	Closes Learnset Manager.
F5	Verifies documents in Supervised Learning.
F7	Manually reclassifies documents. Available to Verifiers and Managers.
F1	Opens Learnset Manager Help.

Learnset Manager toolbar buttons

The Learnset Manager toolbar provides quick access to some frequently used commands.

Button	Description
	Show Settings.
	Switch to Accumulated Documents Browsing (the Local Learnset).
	Switch to Global Learnset Processing.
	Verify Documents. Activates the Advanced Verifier mode.
	Correct tables. Allows you to correct data in the tables. You have to click a table field for this to be active.
	Accept documents.
	Reject documents.
	Learn documents. (Adds them to Global Learnset.)

Learnset Manager viewer toolbar buttons

On the Viewer toolbar, you can use the following commands to adjust the size of a document relative to the width of the Document Viewer window.

Button	Description
	Fits the document to window height.
	Fits the document to window width.
	Best fit.
	Zooms in. Alternatively, press [Ctrl] + [+].
	Zooms out. Alternatively, press [Ctrl] + [-].

Using Learnset Manager

About the Learnset Manager process

Use Learnset Manager to work with Local Learnsets: first, verify the documents, decide whether they belong in the Learnset, add them to the Common Learnset, and train the Learnset.

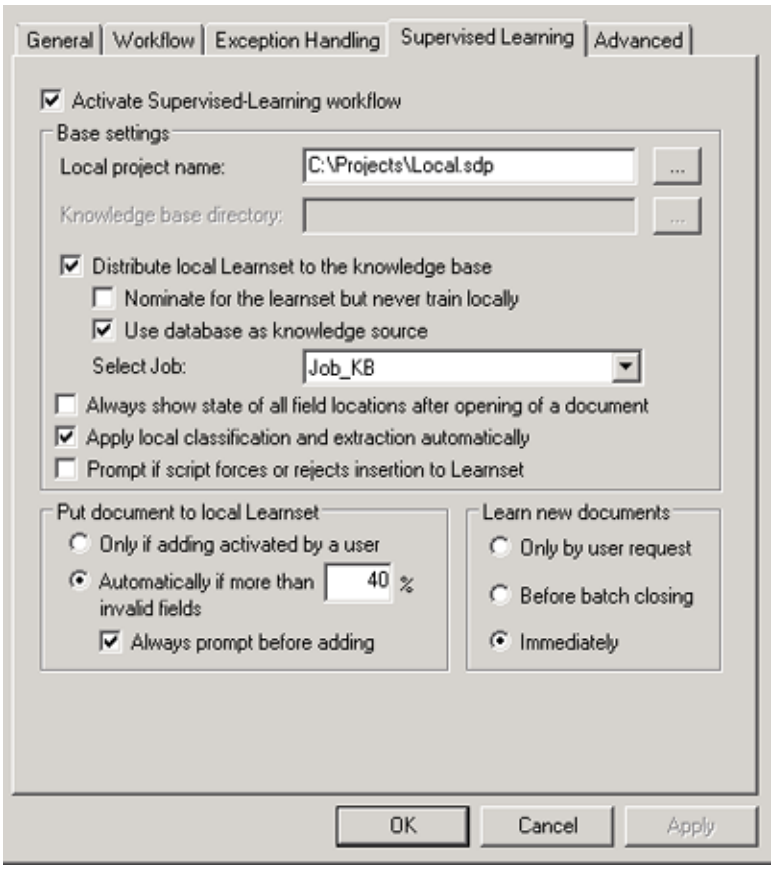
In the Common Learnset, you examine the documents for inclusion in the Global Learnset, accept or reject them, and add them to the Global Learnset. After that, you train the Global Learnset.

Configure Learnset Manager

To configure Learnset Manager, complete the following steps.

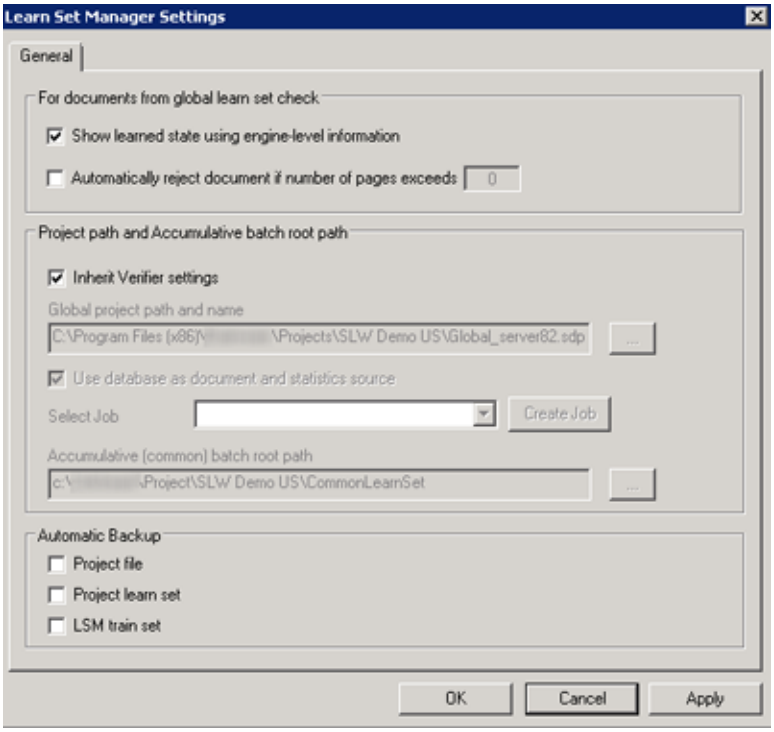
1. In Verifier, examine the properties for Learnset Manager. Most were established in Designer. However, you need to ensure that Learnset Manager is enabled, that the Activate Supervised Learning workflow checkbox is checked. Also, ensure that the paths for Local Project Name, Local Learnset Directory, and Knowledge Base Directory are correct. As a rule, the addresses for each of these base settings should be the same except for the last part.
   - **Local Project Name** should end with the project name (for example, newproject.sdp).
   - **Local Learnset Directory** should end with the Learnset Directory folder (it is probably called Learn).
   - **Knowledge Base Directory** should end with the Common Learnset folder. (it is probably called Common.)

   Note: You can change these settings only if you are an Administrator, Supervised Learning Verifier, or Learnset Manager.



2. Launch the Learnset Manager module from Verifier by clicking the Learnset Manager button on the toolbar.

3. On the Learnset Manager toolbar, click the **Settings** button.



4. Review the settings for document from global Learnset check.
   - **Show learned state using engine-level information**. Make sure this option is checked. This setting indicates whether the particular field or document was used by the system for learning. If required, a user can also disable learning for the desired field/document.
   - **Automatically reject document if number of pages exceeds.** Documents with more pages than specified in the appropriate field will be prevented from being added to the Learnset.
5. Review the settings for Project path and Accumulative Batch Root path.
   - **Inherit Verifier settings**. If this option is selected, the settings made on the Verifier settings tab will be applied, and the options below will be grayed out.
   - If you want Learnset Manager to us a different Global project, or a different job containing data, then clear the **Inherit Verifier settings** option, and populate the options below as to your needs.
   - **Use database as document and statistics source**. Intelligent Capture core information can be stored in the Intelligent Capture database.

   Note: The file system functionality is still supported.

   - **Select Job**. You are able to select the desired job from the Select Job list if you have selected the Database as your source.

   Note: To use the option **Use database as knowledge source** you need to have a job for a Common Learnset in the database. If this Common Learnset job is not available, a database 'job' must be created from RTS with Common Learnset folder as the Batch Root.

- Accumulative (common) Batch Root path. Here you can set the path to the Batch Root. This option is not available when using the Intelligent Capture database.

6. For **Automatic Backup**, select the files you want to be backed up automatically.
   - Project file
   - Project Learnset
   - LSM train set

Work with Common Learnsets

Your first task in Learnset Manager is to attain the highest-quality Learnsets possible. You'll use Learnset Manager to build upon Local Learnsets created earlier in the process. To work with Common Learnsets, complete the following steps.

1. When you launch Supervised Learning from Verifier, the Accumulated Documents Browsing Mode should appear by default. The Accumulated Documents Browsing Mode is the one you use to work with Common Learnsets. The Accumulated Documents Browsing Mode is activated when you click the **Switch to Accumulated Documents Browsing** button or when you select **View** > **Accumulated Documents Browsing**.
2. In the Accumulated Documents Browsing Mode, select a batch to work on.
3. Double-click on a class to select it.
4. Select a document to work on and verify the document just as you would in the traditional Verifier. Click the **Funnel** button on the toolbar to activate the Advanced Verifier Mode where you can correct or verify the contents of each field and table.
5. After you have verified the document, click the **Accept** button. This marks the document for learning as the first step for promotion into the Common Learnset. You could also click the **Reject** button to eliminate the document from being considered for the Common Learnset.
6. Select another document from the batch by accessing the **Learn Statistics Panel** at the bottom of the screen, where you'll double click on a document to open it in the Document Viewer and work on it.
7. When you have verified all the documents you need to verify, click the **Learn** button. This promotes all the accepted documents to the Common Learnset. Notice that the Learn Statistics window for the Local Learnset is now empty.

Correct tables

To correct tables, complete the following step.

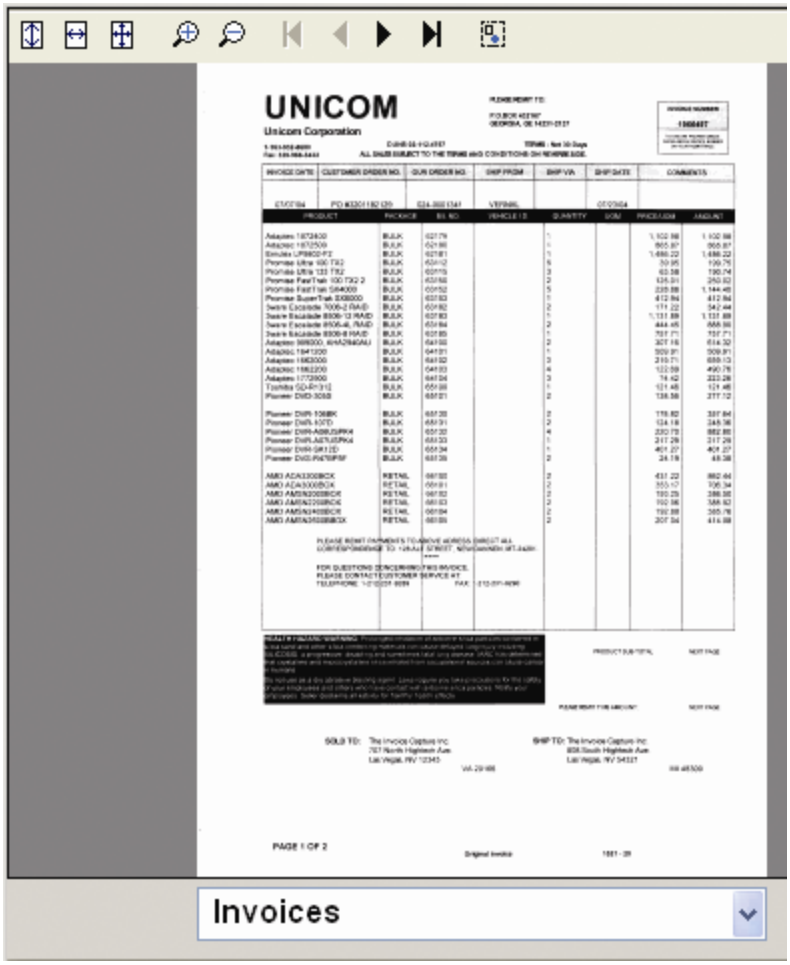
- Select a table in a document and then click the **Correct Tables** button.

Result This enables Supervised Learning Managers and Verifiers to interactively train all the tables on a document form (not just the table you selected so you could activate the Correct Tables button). From there, table correction in Learnset Manager proceeds just as it does in Verifier.

Reclassify a document

To reclassify a document, complete the following steps.

1. To assign a document to a different class, select the **Document** menu from the main menu and select **Reclassify** (or press F7).
2. This opens a dialog box in the Verifier **Document Viewer** where you can assign the document to a new class.
3. When you have selected the new class, press ENTER.



Note: Manual reclassification will only work with the classes that Verifier is currently using, not the classes that have already been learned.

Accept and reject documents

There are two ways to accept or reject a document or batch from a Learnset.

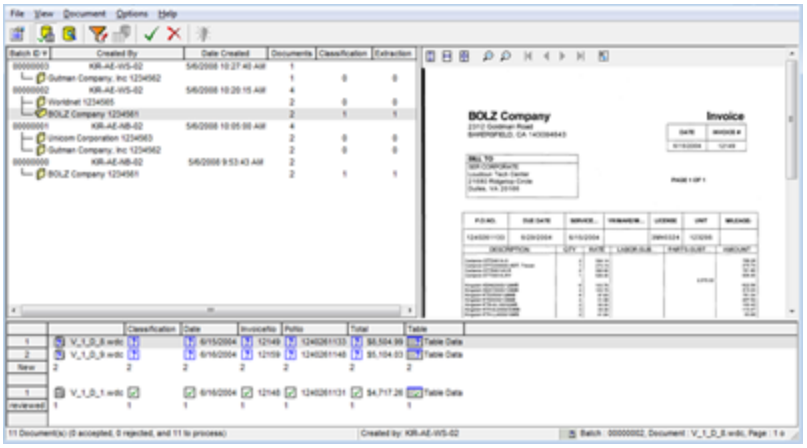
- The first is the traditional way, by using Verifier to manually screen and verify the document or batch.
- The other method is by comparing documents in the Common Learnset to the corresponding batch in the Global Learnset.

Note: If you activate the “Automatically reject document if number of pages exceeds…” checkbox in the Learnset Manager Settings, documents with more pages than specified in the appropriate field will be prevented from being added to the Learnset. The user is notified when a document exceeds the number of allowed pages, and the process continues by choosing Yes in the pop-up window.

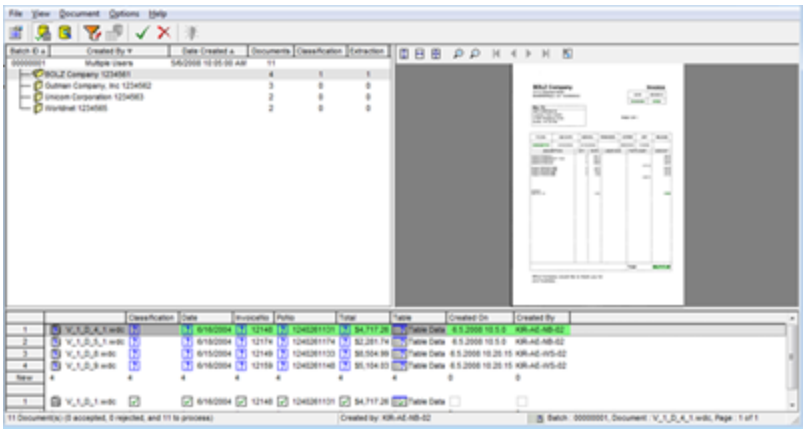
About sorting by vendor and other sorting extensions in Learnset Manager

Learnset Manager can sort by vendor name across multiple batches produced by different local supervised learning Verifiers. This simplifies SLW decisions as to which documents to train for a specific vendor class. With the help of this feature, the user can now review all documents (for the same vendor class) created by different Advanced Verifier workstations at once.

By selecting View > Sort Batches by Vendor, the system rebuilds the batches of documents created through multiple sessions by multiple Advanced Verifier workstations and allow the Learnset Manager user to sort by vendor name. In this connection, each vendor folder cumulates all available documents for this vendor, so that the user could select the best documents to train the global project with.



The Created On and Created By data fields are then displayed separately for each particular document in the Document View at the bottom of the Learnset Manager screen.



Note that in the Global Learnset Browsing mode, Sort batches by Vendor and Sort batches by Date options are disabled. Both of the sorting options are enabled only in Accumulated Document browsing mode. However, in Global Learnset browsing mode, the Vendor classes are sorted in alphabetical order. Furthermore, under the Display Class Name column, class name is shown only for the document classes that already exist in the global project.

Work with Global Learnsets

In the previous section, you learned how to work with Common Learnsets and to promote documents to the Global Learnset. The Global Learnset is where you further refine the quality of your data so that it can be migrated into an effective, useful Global Knowledgebase. To work with Global Learnsets, complete the following steps.

1. Click the **Global Learnset Viewing** button.
2. Again, begin your work at the document level by examining the quality of the data in the document. As before, you select the document from the **Learning Statistics** window at the bottom of the screen. Right-click the document if you are satisfied with it, and then select **Enable**. Select **Use to Train Base Classes**.
3. Click the **Learn** button.
4. Confirm that you do want to learn the document by clicking **Yes**.
5. After you have verified each field in the document, click the **Accept** button on the toolbar to accept the document for processing, or click the **Reject** button to reject it.
6. Now retrain the Learnset by clicking the **Light Bulb** button on the toolbar.

Train base classes

The final milestone in creating or enhancing your Global Learnset is to train base classes. To train a base class, complete the following steps.

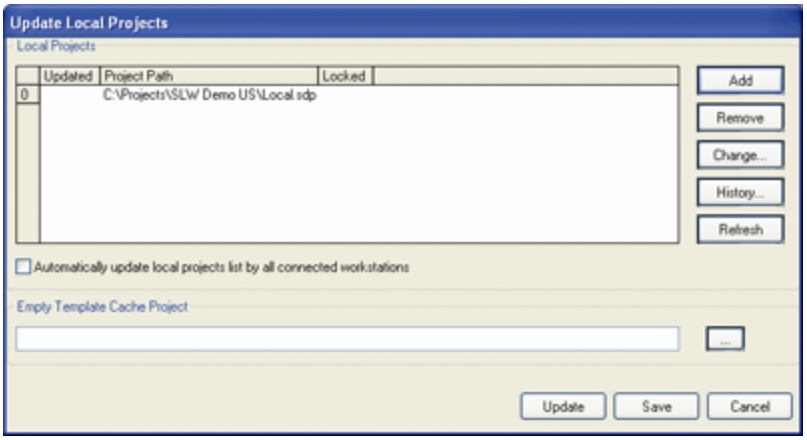
1. On the **Options** menu, select **Train Base Classes**.
2. Select the base class to train.
3. Under **Train Selected Base Classes**, select a value. To avoid errors while maintaining the quality of your sample, select the lowest value possible.
4. Click **OK**.

Update local projects

The ability to update local projects is important for keeping your Learnsets in sync with each other by enabling you to automatically synchronize them. During the work with Verifier, the global project's Learnsets are constantly updated. An administrator may then wish to update the local projects (which are out of date) with a new global project. They use this functionality to add all the local projects they want to update into the main list. Then they point to the global project template (the project that will be used to overwrite all the out-of-date local projects), and press update to update them. To update local projects, complete the following steps.

1. On the **Options** menu, click **Update Local Projects**.
2. Configure the selections described below and click **Update**. This procedure can take a while, especially if you are updating projects on a network. Note that locked projects will not be updated. The Update Local Projects dialog box specifies a list of local or network project paths to be managed. Here, you can:

- Add project paths to the Update list by clicking **Add** and browsing to the project.
- Remove paths from the Update list by clicking **Remove** and browsing to the project.
- Change existing project paths by clicking **Change** and browsing to the project.
- View a history of all Verifier workstations that have connected to the Common Learnset. The History shows the workstation name, the time, and date of its last connection and the local project path.
- This list is updated every time a Verifier station creates a new batch of locally learned documents in the Common Learnset.
- Refresh the list of projects to see the most recent update information about them.
- Empty the template cache.
- The **Empty Template Cache Project** field is used to update the local project.
- Update the list of projects or save the new criteria without actually updating the projects.



Result For each configured path, the dialog box shows whether a project is up to date, whether it is locked, and whether the project is available. An up-to-date project has a green check mark beside it; a project that has not been updated has a red X. Path names of unavailable projects are dimmed. Note that the settings you establish above will be available for any workstation on which Learnset Manager is opened.

About using Learnset Manager on several workstations

Learnset Manager can be used simultaneously on more than one workstation thanks to the application's ability to lock projects and files.

Learnset Manager supports three levels of protection that facilitate this ability.

- Batch-level locking.
- Allowing all Learnset Manager workstations to view changes made by all Supervised Learning Managers.
- Locking project files and Learnsets while they are being trained.

About batch-level locking

Batches are locked when they are in process; it prevents several users from updating the same batch at the same time. No one else can access the batch until processing is completed and the batch is closed.

About tracking changes made by Supervised Learning Managers

The changes applied by one Learnset Manager user should be visible to all other Learnset Manager users. To accomplish this, Learnset Managers must use two predefined batch document states: “981, accepted” and “982, rejected”.

When learning is executed, the Learnset Manager application checks to see if documents with either of these assigned states have been added to the Local Learnset. Documents with a state of 981 are added to the Global Learnset. Documents with a state of 982 are not added to the Global Learnset.

About project file and Learnset locking during training

Only one workstation can do learning at one time. This means that the learning process is locked, and therefore not available for other users, if one user has initiated learning.

Tips for tricky situations

The tips in this section can help to resolve some situations during the extraction and validation process.

My document contains an invalid extraction result. However, this result is precisely what I need, and I want to validate the field. What can I do?

This depends a bit on the design of your application. In most cases, you will have to press ENTER three times.

In one of my batches, there is a document that must be classified manually, but it does not belong to one of the available classes. I cannot release the batch as it is. What can I do to finish my job?

Normally, your organization will have specialized workstations where people are in charge of handling special cases that only occur as exceptions.

In one of my batches, there is a document I have already validated. However, I’ve overlooked a mistake in this document. I don’t want to release the batch without correcting it.

You can use the Document Mode to get to the document. Select the document and switch to Verify Mode. Make corrections and press ENTER.

Sometimes the indexing window looks odd: It has no field area, only the current input area. How do I get to the next field?

Not a problem. You can use all keyboard shortcuts for field navigation from within the current input area.

When I switch from one field to the next, the document is not moving as well. I find this annoying. Is there a way to stop that?

Yes, there is. With your current settings, the application always searches the document area associated with the current field’s content. This area is then displayed. To turn this off, click on keep focus: Alternatively, you could just use a different magnification ratio.

I want to start Learnset Manager, but I don’t want to have to go through Verifier first. Can I do it?

In a word, no. Learnset Manager is an add-in that can only be started in Verifier.

OK, I tried to start Learnset Manager in Verifier, and I still can't do it. Why?

There are three main reasons.

- You may not have permission to use this add-in. Check with your project administrator to see if you are assigned to a group that can work with Learnset Manager mode.
- Learnset Manager might not be enabled for the project. Again, contact your project administrator.
- A third reason might be that Learnset Manager mode is not properly licensed. Learnset Manager gets its license through a Runtime Server process. If you've been able to get into to Learnset Manager mode before, but you can't now, it may be that Runtime Server has stopped. If restarting Runtime Server does not fix the problem, your project administrator should contact Customer Support.

I want to change the default colors, background colors and fonts for Elements of Verification forms. Can I do it?

Intelligent Capture supports a set of script methods to dynamically or statically adjust fonts, colors, and background colors for verification forms and their verification elements. Refer to the Scripting Guide documentation for more details.

Appendix: Quick reference

This appendix provides a quick reference to some of the commonly used toolbar buttons.

Main Controls

Icon	Control
	Batch View
	Start Verification
	Batch Structure
	Display Properties
	Display Batch Filter Dialog
	Start Learnset Manager
	First Batch Page
	Previous Batch Page
	Next Batch Page
	Last batch Page
	Refresh

Verification View

Icon	Control
	Exception State
	Highlight Candidates
	Keep Focus on Field
	Keep Zoom

Icon	Control
	Previous Page in Document
	Next Page in Document
	Rotate Image (90 degrees)
	First Document in Batch
	Previous Document in Batch
	Next Document in Batch
	Last Document in Batch
	Correct Tables
	Add Current Document to Local Learnset
	Classify and Analyze Current Document

Document View

Icon	Control
	Fit to Height
	Fit to Width
	Fit to Size
	Zoom In
	Zoom Out
	Single Page View
	Two Pages View horizontally
	Three Pages View horizontally

Icon	Control
	Two Pages View vertically
	Move Image Downwards
	Move Image to Left
	Move Image to Right
	Move Image Upwards

Learnset Manager

Icon	Control
	Show Properties
	Accumulated Documents Browsing
	Global Learnset Browsing
	Verify Document
	Accept Document
	Reject Document
	Learn Documents

Glossary

Accumulative Learnset

The common Learnset.

Administrator

In Intelligent Capture, an administrator is a power user who creates user accounts, passwords, and groups, and assigns users to groups.

Analysis

In this processing step, the document content is analyzed and a set of possible values for a field is generated. These values are called candidates.

Associative Search Engine

Uses a reference field to extract results.

Automatic Supervised Learning

Uses the Associative Search Engine to process, classify, and extract information.

Base class

The highest level of a classification.

Batch

A logical organizational structure to control a set of documents during a process. A batch is normally created during the scan process from a batch of paper. The status of a batch is used to manage the input flow.

Brainware Table Extraction

An extraction method that facilitates interactive table training.

Candidate

Set of possible values for a field.

Child class

A class spawned by a parent class. See also base class and parent class. Also called a sub-class.

Class

A set of documents that are grouped by common content. Each class usually has a mnemonic name that describes its contents from the user's point of view.

Classification

The process of assigning one or more classes and corresponding confidence values to one or more unknown documents.

Common Learnset

An accumulation of Local Learnsets.

DocClass.

A parent document class.

Document

Any electronic file mainly consisting of ASCII text. If this is initially the case, OCR or filtering must be applied to create the text representation. A document can be classified, have fields used for extraction, and have one or more images attached.

DPI

Dots per inch. Affects the size and clarity of an image file.

Evaluation

The process of determining a class or the contents of a field from confidence levels, weights, or distances for classes or candidates.

Export

In Intelligent Capture, document export releases the documents so that they are no longer managed by the software.

Extraction

The process of automatically finding specified information within a document and writing the information to data fields associated with the document. Extraction is used for automatic indexing.

Folder

A logical structure inside a batch for coherent documents. For example, a folder may consist of all pages of a correspondence with many folders inside one batch.

Form

(1) A structured, standardized document that is used to support business processes.

(2) A custom dialog box in a software application.

Global Learnset

A general Learnset that encompasses similar classes or projects. See also Local Learnset.

Importing

Bringing documents into Intelligent Capture for management and processing.

Indexing

The process of assigning attributes to a document. This can either be done manually, semi-automatically (Smart Indexing), or entirely automatically (Extraction).

Intelligent Capture

Intelligent Capture analyzes text from any media type. It uses artificial neural network techniques to automatically classify structured and unstructured documents and extract meaningful information from them. A neural network must be trained before you can use its ability to categorize at high speed – it must learn. The method of learning is similar to the way humans learn: it is purely sample-based. The major benefit is that, after being trained, Intelligent Capture can handle information that is similar to the samples without programming or extensive rule setting. Intelligent Capture can operate at high speed and can be implemented on parallel hardware to further enhance performance.

Knowledge Base

A database of knowledge about a subject; used in artificial intelligence. The knowledge base comes partly from human experience and partly from the computer's experience.

Learnset

In classification, a Learnset is a set of documents whose class assignments are specified by the user. For each view and each class, the user must provide a sufficient number of representative documents. Similarly, in extraction, a Learnset is a set of documents whose field contents are selected by the user from a set of candidates.

Learnset Manager

A user who designs, modifies, and maintains Learnsets.

Learning

Given a view with a set of documents in vector representation and their class assignments, a neural network is created, so that the defined classes can be reproduced without error. This neural network is then used in all subsequent classification tasks.

Literal character

Normal alphanumeric characters that are not used as operators.

Local Learnset

Learnset specific to a document class.

Neural network

An artificial neural network is an application that in some ways works like a human brain. This includes the ability to learn. It consists of artificial neurons that are linked into a network of layers. The neural network can receive signals through an input layer, process it within the internal layers, and send signals through the output layer. During learning, a specified input (called a teacher signal, such as documents from a Learnset) and the desired output (such as the corresponding classes) are presented to the network together. Processing is then adjusted until the desired output can be produced from the teacher signal.

OCR

Optical Character Recognition. The reading and recognition of symbols of text from a piece of paper or a scanned image. OCR detects the symbols and converts them into characters and words that can be read electronically.

Parent class

A class with derived classes, called children.

Persistent

Permanent; something that is saved persistently is saved permanently, unless a user or process deletes it.

Project

Project files are used to persistently save custom settings for Intelligent Capture applications. They are created in Verifier and handed over to Runtime Server for productive operation.

Smart Indexing

Smart indexing uses a database lookup to determine document attributes. It can be used for automatic indexing and to support manual indexing.

Sub-class

A derivative class. Also called a child class.

Supervised Learning Verifier

A user who collects and maintains local training data.

Validation

A quality assurance task that involves confirming whether a processing result is correct. This can be done at several levels: for the class or a field associated with a document, for the document as a whole or for an entire batch.

Verification

A quality assurance task that involves checking and correcting processing results.

Verifier

The Intelligent Capture QA application.

View

A set of documents that represent at least two classes. A view is usually defined using a small set of documents that represent the domain of interest. In a view, classes compete for documents; that is, a document may only be assigned to one class within the view.

Web Verifier

The Intelligent Capture web-based extension of the Verifier Thick Client.

Workdoc

An internal structure representing the logical structure of a document. The Workdoc represents the data created during processing of a single document and is stored in a file with the extension *.wdc. Since the Workdoc includes all OCR and analysis results, it may exceed the document file by size.