Perceptive Intelligent Capture Web 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 is Web Verifier?

Perceptive Web Verifier is a web-based extension of the Verifier Thick Client. It is an application module to allow users to verify documents with no software installed on the client side. The processing of activities primarily occurs on a central server allowing use of thin client devices running only a web browser.

This is an outstanding advantage for your business: keeping only a few servers busy and the workstations running the Web Verifier results in easier system management and lower costs. Furthermore, this solution allows for central management of storage and backup and provides for easier security.

Thus, the Web Verifier solution will contribute to considerable reduction of the cost- and personnel-intensive maintenance of several workstations.

Another significant advantage is higher flexibility. As the only thing you need to work on your project is a workstation running a web browser and an Internet connection, you are able to access your project or the current batch list from everywhere. Furthermore, the system is able to remember the application state after the user's last session. When the user starts the application next time, the system will load the last application state. As this information is stored in the database on the central server (as user-level project settings), access to it through authentication using your login information from every device irrespective of the operating system is possible.

In addition, users of the Web Verifier module have the option to apply lockers to resources for exclusive use. Thus, no other module or user will be able to use this resource. However, in order to prevent dead locks, each locker is provided with a timeout. This means that upon exiting the application, the system will remove all user owned lockers.

These are some key features of Web Verifier.

- 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 Web 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.
- Through automatic locking, document batches can safely be processed by teams of operators.

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.
- The Intelligent Capture application 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.

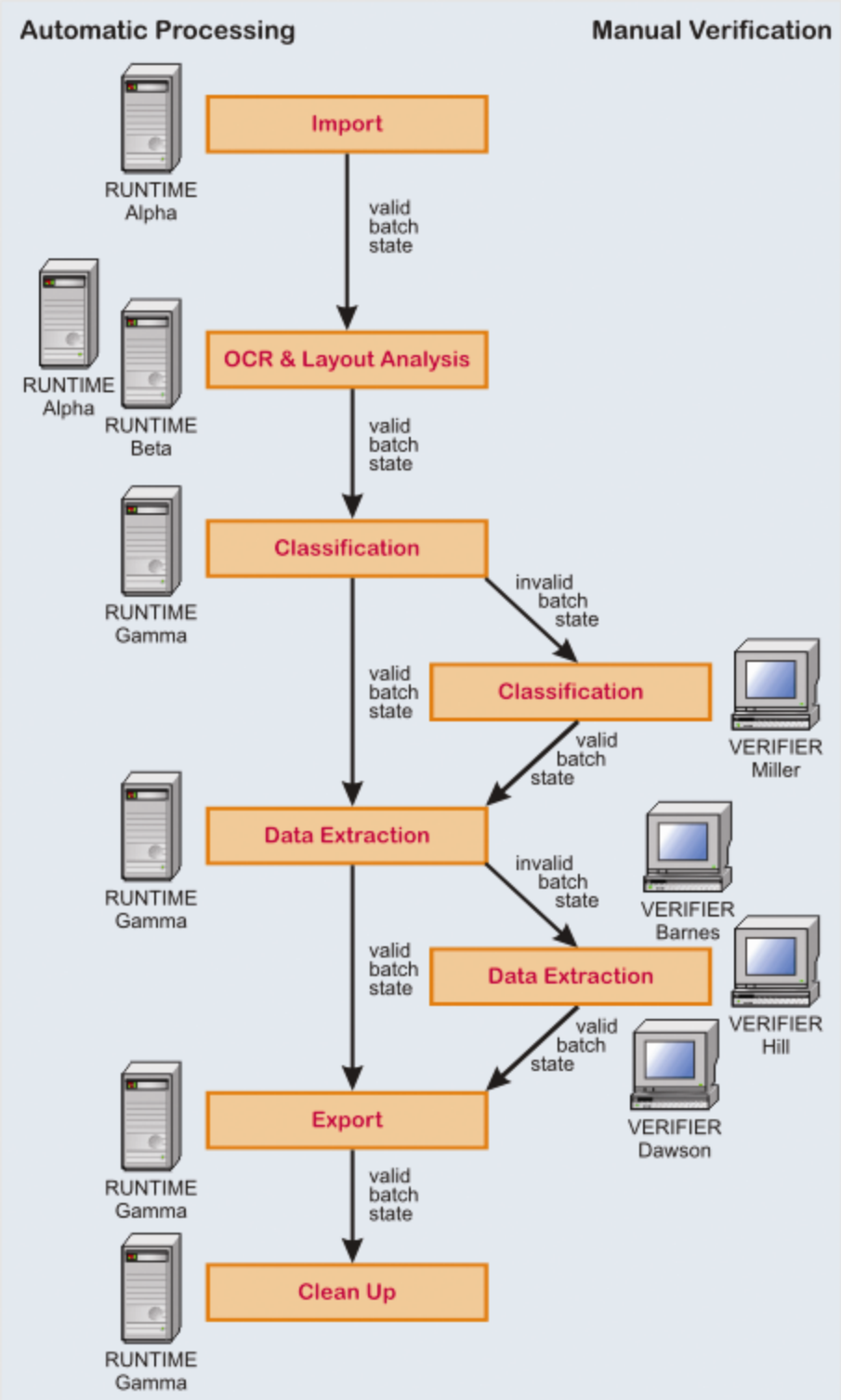
However, if the Runtime Server detects that the batch contains invalid results, the batch must be routed to a verification station, where you can analyze and resolve the problem using Verifier. Invalid batches are presented to you in a task list, called Batch View. You will have to resolve each problem 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 the Intelligent Capture application 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 shows 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
RUNTIME
Gamma
VERIFIER
Dawson
valid
batch
state
Clean Up
RUNTIME
Gamma

About installation

If your workstation is intended to run only Web Verifier, no installation steps at your workstation are necessary. All you need is a web browser, an Internet connection and your login information.

Your administrator will create a user account for you for the Intelligent Capture context using the Designer application, and will assign at least one role to you.

In context of the Web Verifier, all work associated with installation, user assignment, project design, and database creation is done at a central server. This should be done by an experienced user.

About language support

The Web Verifier application implements full support of CJKT. Among other advantages this means that the user can review and edit and validate the extraction results using multiple non-Western languages at a time.

Searching with the Associative Search engine, for example, can be applied to a pool that contains multi-language entries. In this case, the user can even search for a mixed text with a couple of non-Western words or characters typed together within one search query, but in this case the “hieroglyphic” pool items may obtain better weighting compared with the other languages. This is however specific for Associative Search Intelligent Capture V3 based processing.

This applies to the following workflow steps in Intelligent Capture.

- Classification
- Extraction
- Validation
- Learning

Start and exit Web Verifier

About starting Web Verifier

If Web Verifier is installed as recommended, you can launch it from your web browser using the server or application address. After you log in, the application displays a window where you can select a project and the job.

If you start Web Verifier on a machine for the first time, all of the controls may not display. If this is the case, you have to first add the Web Verifier site to the trusted sites list of your Internet browser.

Note: The Sax Basic scripting engine is no longer supported. If your application is configured to open a project that 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)`. In this scenario, open the project in the Designer application and migrate it to use the WinWrap scripting engine.

About logging into Web Verifier

When you log in to an existing project in Web Verifier, supply your Intelligent Capture user name and password. Your user name and password are assigned to you by your project administrator.

After you first log in, a dialog box may open for a password change if your administrator configured this requirement. If this is the case, you must change your password before using the application.

The user name and password are used for the login to the Web Verifier platform. The user name and password also set the user rights for the active session. The name of the current user is displayed in the top-right corner of the Web Verifier window.

Note: If you have questions or problems with your user name or password, please 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 logins. This has been enabled if the **Remember me next time** check box appears on the log-in form. To remember your user name and password between logins, complete your user name and password and select **Remember me next time** before clicking **Log In**. The next time you log on with the same computer, the system will complete the user name and password automatically. You also have the option to select the language for the application from the list in the bottom-right corner.

Log in to Web Verifier

To log into Web Verifier, complete the following step.

- Start an Internet browser and open the Login page of Web Verifier. This address is provided by the project administrator.

  Note: The login for Web Verifier is configured by the project administrator using either the Web Verifier login or Windows authentication.

About Windows authentication

In order to use Windows authentication, your administrator needs to configure the Web Verifier service to allow Windows authentication.

The Windows authentication login for Web Verifier shows the following special behavior.

- A canceled or failed authentication opens an error page. Press the reload button to open the Login window again.
- You can switch to another user, such as the Administrator, by opening the File menu and clicking Re-login. The standard Web Verifier Login window opens.
- Windows authentication mode prevents changing your password.
- The Logout link terminates the session on the server. Reloading a Web Verifier page in the same browser instance will create a new session on the server for the current user.
- In order to completely logout from the system, it is necessary to close all open browser sessions. Otherwise, returning to the starting page would automatically restore the last session. A message screen at the end of the logout step reminds you of this:
  `Thank you for using Web Verifier. To complete the logout, you may now exit all opened browser windows.`

Log in using Windows authentication

To log in using Windows authentication, complete the following steps.

1. Open the start page of the Web Verifier application.

   If the current computer user is configured for Web Verifier Windows authentication by the project administrator, the Web Verifier interface opens. The Windows authentication form may display.
2. Complete your Windows credentials if the Windows authentication form appears.
3. When logged in with Windows authentication, the Web Verifier window shows the domain and computer name in the top-right corner.

Issues you may encounter during login

This section covers issues that you may encounter at login.

Maximum number of users exceeded

There is a maximum number of users allowed to access the Intelligent Capture database at a time, which is defined by the license. If the maximum number of users is exceeded, the following message displays (and varies depending on the number of licenses available).

```
The system is unable to log in, because all five user licenses are
currently in use. Please contact Customer Support for further assistance.
```

Session timeout

To maintain the workflow with several Verifier users, a Web Verifier session ends automatically after 20 minutes of no activity. All batches are released and lose their changes that were applied during the last session. Any action in Web Verifier after a timeout opens the login screen with the following information.

```
Your session is currently expired due to either a log out, or a timeout
of 20 minutes. Please log on once more.
```

Exit Web Verifier

To exit Web Verifier, complete the following steps.

1. Click **Logout** in the top-right corner.
2. Close your web browser.

Users, groups, and roles

To load a project, you must log in with your user name and password. The user name and password combination not only lets you into a project, it governs what you can do once you get there. User name and password combinations are set up by your project administrator.

Your administrator also sets up user groups and assigns you to one or more of these groups. In turn, the administrator assigns one or more roles to each user group. The user role is always more prominent than group role.

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 Intelligent CaptureVerifier configuration.

Verifier Filtering

The FLT role is to allow a 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 login password

To change your password, complete the following steps. For users logged in with Windows authentication, the Change Password option is not available.

1. Select **Change Password** from the **Options** menu.
2. Type in your existing password in the **Old Password** box.
3. Enter a new password.
4. Re-enter the new password for verification.
5. Click **OK**.

About configuring Web Verifier

Configuring Web Verifier requires specifying which batches of documents are processed at a given station.

This includes:

- The Designer project file that contains the settings used to process the documents. The name of the job that contains the batches to be verified.
- The processing steps that you want to verify: classification, extraction, document separation, or all of them.
- The status of batches before and after processing.

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

Open the Properties dialog box

The Properties dialog box is where you configure Web Verifier. To open the Properties dialog box, complete the following step.

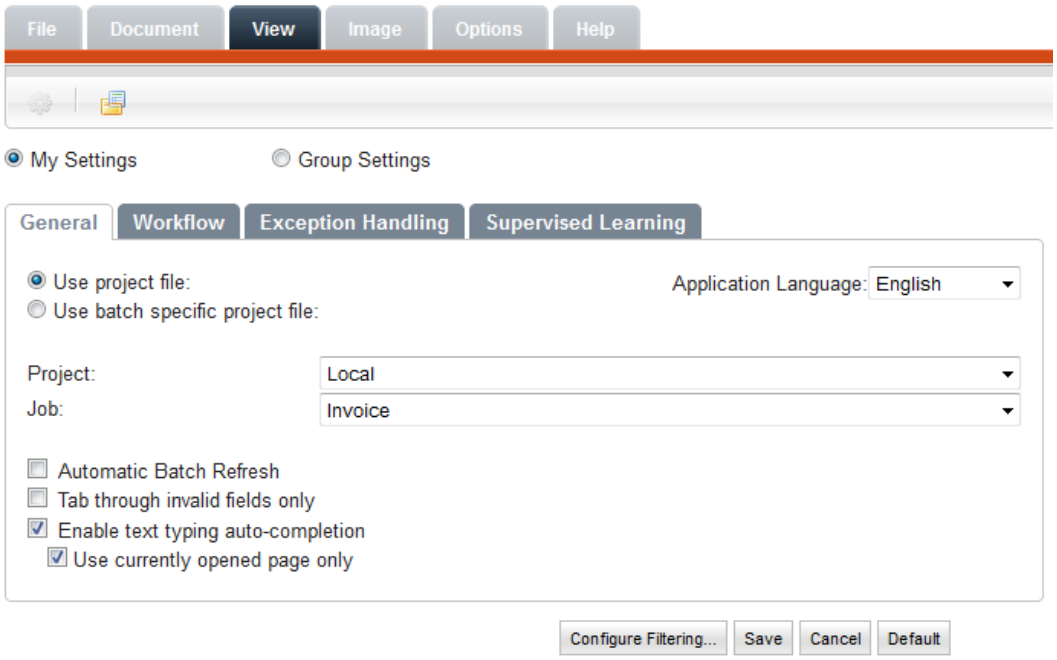
- Click the **Settings** button in the toolbar.

General settings

For general settings such as referenced project and job, select the General tab. Web Verifier presents the My Settings view. When the Administrator creates users, he usually assigns them to one of the available groups with specific roles assigned. Depending on the group the current user belongs to, the following Settings views are possible.

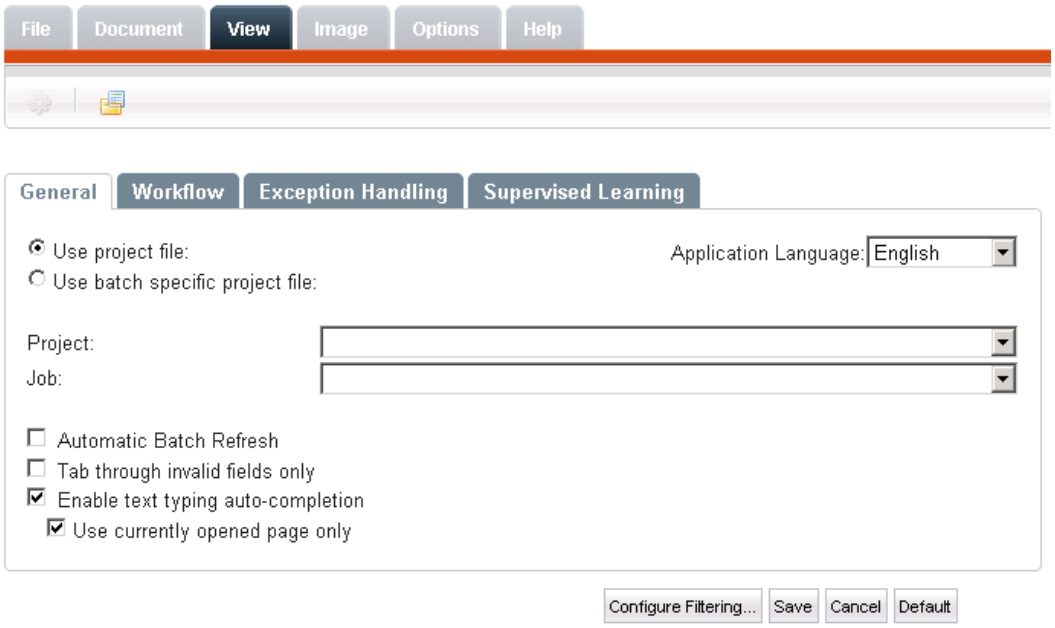
Administrator user

For a default Administrator user with all of the available roles assigned, the Settings window looks as shown below.



SLM and SLV roles

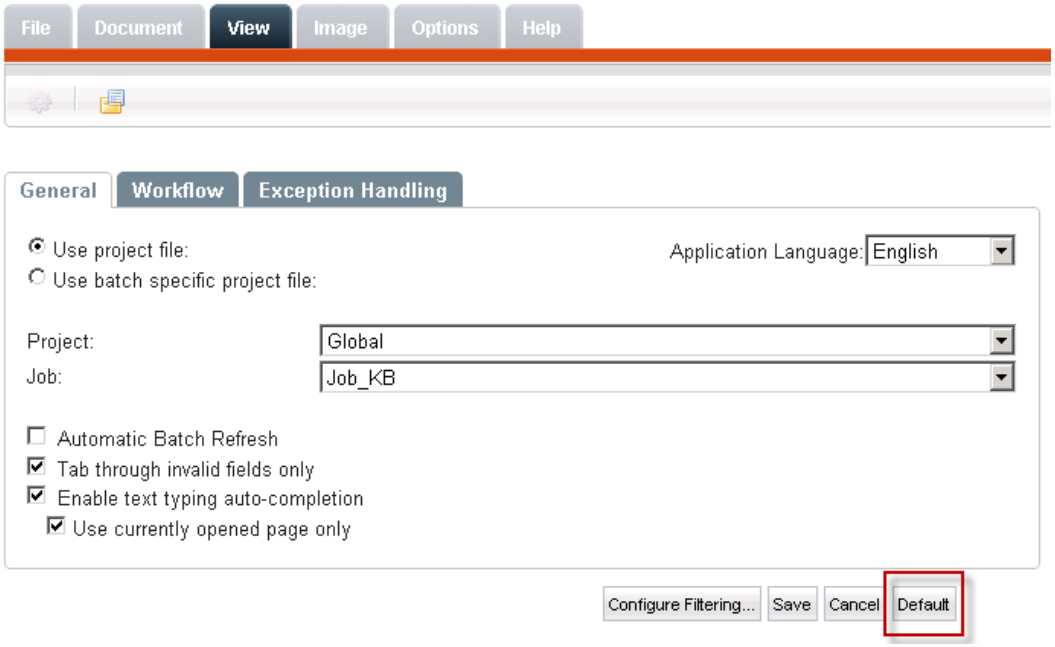
Learnset Manager users with the SLM (Supervised Learning Manager) and SLV (Supervised Learning Verifier) roles assigned, but without the ADM (Administrator) role are presented with a Settings view as shown below. The Group Settings are not accessible, but the Supervised Learning tab is available.



Verifier user

For a default Verifier user, with the VER (Verifier), SET (Verifier Settings), and FLT (Filtering) roles assigned, the Settings window will look as shown below, after clicking the Default button. Usually, the Administrator has already configured the Group settings with the project selected. Clicking the Default button adopts all settings for your group. The fields in all available tabs populate automatically.

No Supervised Learning tab is available, and you are not able to edit Group Settings.



If you want to configure your own settings, please find the descriptions for the single configuration options as indicated in the following sections.

About specifying the project file

The **Use project file** option is for selecting the file name of the Intelligent Capture project used to process the documents; this contains the design of the indexing form that you will use to verify the extraction. The following scenarios are possible for the project loading procedure.

- When you log in to Web Verifier for the first time, and the project is already configured for you, the system prompts you to let you know that the project is loading.
- When you select a new project and click **Save** to save the settings made, the selected project loads after you returned to the Batch View and click **Show Batches**.
- If you are using the **Use batch specific project file** option, the project loads as soon as you attempt to open a batch, and each time a batch is opened for a different project.

About specifying the application language

The application language can be changed in the General settings.

Web Verifier supports an extended list of languages: Chinese Simplified, Chinese Traditional, Danish, Dutch, English, French, Finnish, German, Italian, Japanese, Korean, Norwegian, Polish, Portuguese, Romanian, Russian, Spanish, Swedish, Turkish.

Note: By default, the Chinese language selection points to Traditional Chinese. In case Simplified Chinese language is required, please contact your system administrator.

About turning on automatic batch refresh

If you select the **Automatic Batch Refresh** option, the Batch View automatically shows newly generated batches with matching states. If you do not want the automatic update, you can clear the check box. This leaves you the option to refresh the Batch View (click **Refresh** under the Batch list) when you need up-to-date information.

About specifying tabbing behavior

You can customize the system behavior when a user presses `TAB`.

If you select the **Tab through invalid fields only** option, when the user presses `TAB` in Document Verification mode, the system tabs through invalid fields only. Similarly, when the user presses `TAB` inside a table control, the system tabs through invalid table cells only. This option also applies for tabbing in backwards direction when the user presses `SHIFT + TAB`.

About enabling auto-completion

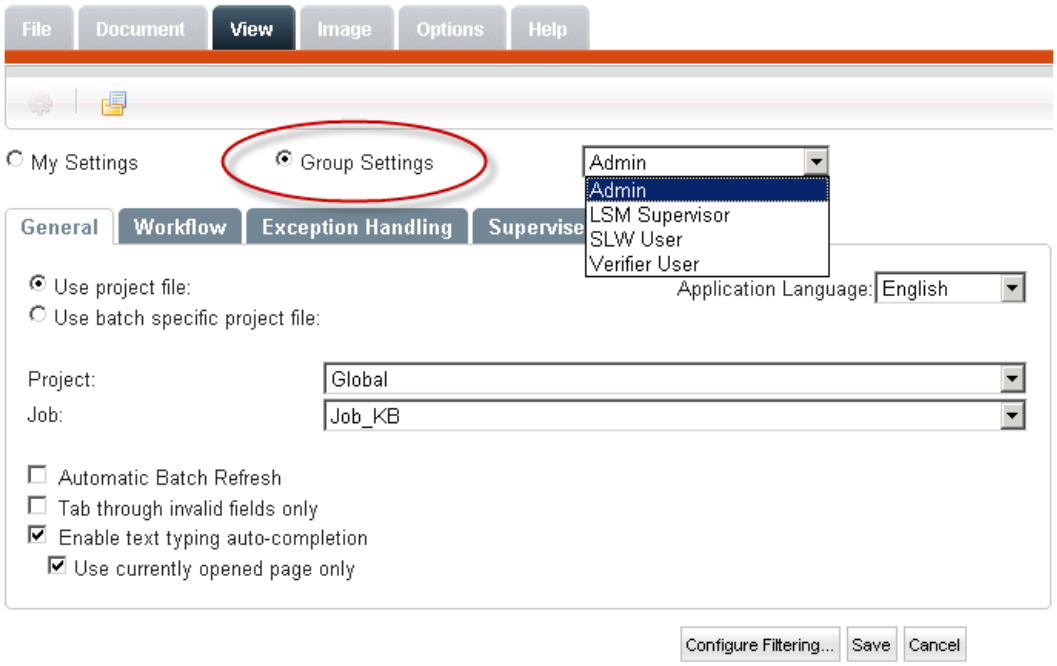
You can enable the **Enable text typing auto-completion** option to speed up typing. When you start to type, auto-text completes the word, suggesting the best match among all of the words or candidates available after OCR and Format Analysis. For example, you can type the first two characters of a 20-character invoice.

You can also enable this function for **Use currently opened page only**; this check box is only available when the auto-completion option is checked.

About group settings

This part of settings is only available for users who have the ADM (Administrator) role assigned. Other users configure their own settings under My Settings. The Group Settings option will not be accessible for them.

The Administrator can configure settings of different groups in here by choosing the specific group from the drop-down list.



About the workflow settings

Runtime Server assigns predefined 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.

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.

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

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

- Manual indexing or extraction correction after manual classification. Note that if a document can not be classified, automatic extraction probably will not work.
- Export after manual indexing or 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 to be carried out at the current Web Verifier station, select the Workflow tab, and then choose from among the following options.

- To configure classification verification, extraction verification, or document separation, select one of the following options in the list.
  - Classification verification
  - Extraction verification
  - Document separation

  Note: One or more of these steps can be performed at the workstation.
- To add an input value, right-click on the **Input** box and select **Add State** on the shortcut menu.

  Note: You can also change states and remove states this way.
- To set an output value, select it from the **Output** box.
- To force Intelligent Capture to attempt to automatically extract data after you manually classify the document, select **Perform automatic extraction after manual classification**.

  Note: 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.
- To skip validated documents and show the next invalid document when you click the **Next Document** button, select **Disable navigation to valid documents**.
- To go directly to the next invalid batch without the confirmation release dialog, select **Hide batch release dialog**.
- To determine the next invalid batch to be presented without prompting, select **When entering document verification mode, by default open next available invalid batch**.

  Note: When this option is selected, the next batch number will be presented. When this option is cleared, the first batch at the top of the queue will be presented.

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 CaptureVerifier 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. To select a state, on the **Exception Handling** tab, mark the corresponding check box.

   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

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 the 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.

Configuring Supervised Learning

Supervised Learning allows users to nominate documents to the knowledge base job. Only users with the SLV role assigned will be able to access this Settings tab.

Activate Supervised Learning workflow

Supervised Learning workflow allows the user to copy documents from current job to knowledge base job.

Knowledge base job

The job name which is used to keep track of the documents added to the knowledge base.

Always automatically re-extract data for poorly processed documents

Global extraction will be applied on documents with low extraction rate.

Copy document to global knowledge base

Only if activated by the user. Copying of the document to the global knowledge base has to be triggered by the user manually by clicking in the Document View page.

Automatically if unclassified or more than... % of fields are invalid. Documents will be copied automatically to the global knowledge base if they meet the configured criteria.

Note: If **Only if activated by the user** is not selected, and the document does not meet the criteria to be copied automatically, you can still copy a document to global knowledge base by clicking in the Document View page.

About batch filtering

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

The filtering dialog box is accessible outside the settings dialog box so that users without a SET role but with the FLT role are able to filter batches. Only users who have the FLT role assigned are 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 your filter settings, complete the following steps.

1. Select Filtering from the list under **Options**, or click on the **Batch filter** icon. Users with the SET role assigned can also access the filtering window from within the **Settings** dialog window.
2. In the filtering configuration window, 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.
3. On the status bar, the currently applied filter conditions are displayed. You can clear the filter condition settings by clicking on **Clear Condition** within the filtering dialog. The filtering conditions are combined with the 'Show batches' condition using logical AND.

4. Within the **Group Settings** section, administrators can easily define filter conditions for several user groups.
   1. Select a group from the drop-down list.
   2. Click **Configure Filtering**.
   3. Establish the filtering conditions and confirm by clicking **Apply**.
   4. Click **Save** to save your settings.

   Users assigned to this group will now be presented with a subset of batches matching the filter conditions.

Result In Web Verifier, the filtering settings are saved for the next sessions.

Getting familiar with the user interface

Access batch view

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

- 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 shortcut keys

The following shortcuts are available in Batch View.

Keyboard Shortcut	Command
`CTRL + 0`	Print

Batch view toolbar buttons

The toolbar provides quick access to some frequently used commands.

Button	Description
	Display a property sheet where you can configure Web Verifier.
	Display a dialog box where you can configure the batch filtering conditions properties.

Button	Description
	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
	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. • Open selected batch • Open the next 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.

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

You can sort a batch list by each column. The table columns display the following information about a batch.

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

Batch ID

A number you can use 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 batch is processed, 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. This field is optional.

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.

Last user/Module

The computer name of the operator who 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 that is assigned to a batch relates to security. Batches can be assigned to a user group using a unique ID.

External Batch ID

The name of the Batch Group. You can use this ID to represent any piece of information you would like to associate with a batch, such as an external system ID or a 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. You can use this ID to identify originators of batches 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. You can use this to identify the types of documents, such as invoices and claim forms, in batches or in a source of document such as an email or a scanned document.

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 Batch View, complete the following step.

- In the table of batches, click on the batch you want to select.

  Depending on the status of the batch, either the classification or the extraction verification view opens.

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 + 0	Print
CTRL + ALT + HOME	First Document

Keyboard Shortcut	Command
CTRL + ALT + PAGE DOWN	Previous Document
CTRL + ALT + PAGE UP	Next Document
CTRL + ALT + END	Last Document
CTRL + 8	Append Document
CTRL + 9	Cut Document
CTRL + ENTER	Accept/Reject next unsure page
CTRL + SPACE	Select next unsure page
CTRL + 1	Show batches
SHIFT + ADD	Zoom in
SHIFT + 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 + H	Fit to Height
CTRL + W	Fit to Width

Keyboard Shortcut	Command
CTRL + F	Best Fit
CTRL + J	Increase image area
CTRL + K	Decrease image area

Main toolbar buttons

The toolbar provides quick access to some frequently used commands.

Button	Description
	Not available in document view.
	Displays the batch view.
	Starts 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 open the selected batch or open the next invalid batch.
	Displays 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
	Displays the first page of a document. This button is only enabled if the current document has more than one page.
	Displays the previous page in document.
	Displays the next page in document.
	Displays the last page in document. This button is only enabled if the current document has more than one page.

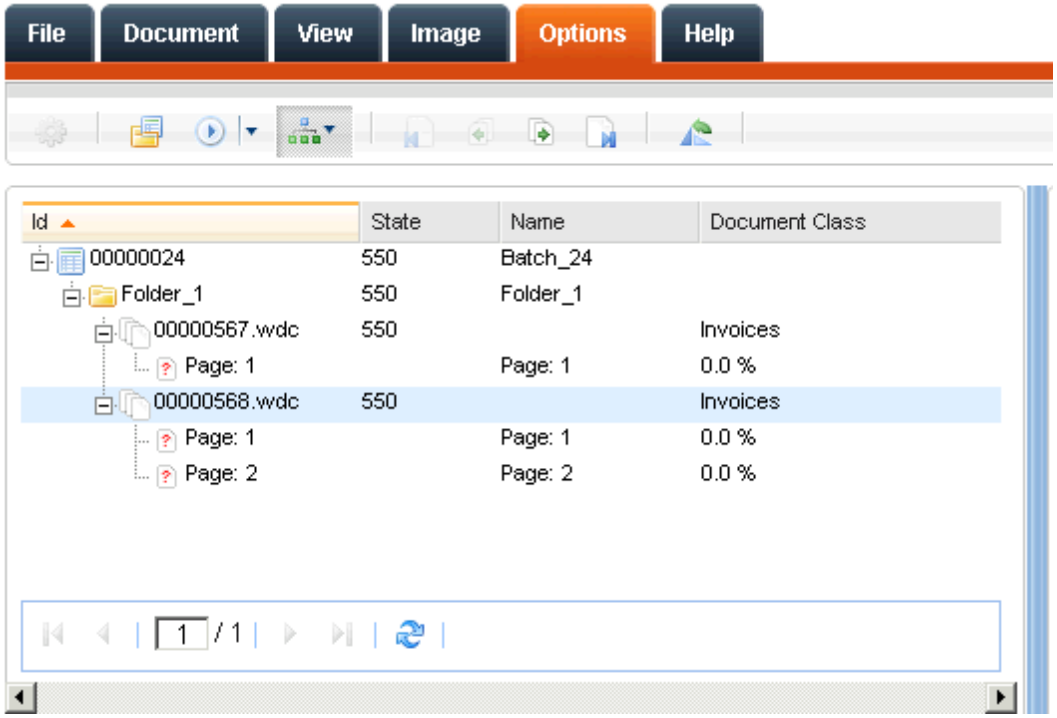
Button	Description
	Rotates the current page clockwise.
	Shows a single page.
	Shows two pages horizontally.
	Shows three pages horizontally.
	Shows two pages vertically.
	Displays the first document in the batch.
	Displays the previous document in the batch.
	Displays the next document in the batch.
	Displays the last document in the batch.

About the batch structure area

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

The levels of this hierarchy are:

- Batch
- Folder
- Document.



For each document entry, Web 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.

Note: Depending on your access rights, the document view might list documents that are inaccessible. This is due to their input states which are out of the workflow you have access to.

Navigate in the document view

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

- To select a document in the batch structure, click on it.
- 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**, or click **Show Selected Batch**.
2. In the document list, click on the desired multi-page document. This document must have at least two pages.
3. Right-click on the page.
4. Select **Cut pages into a new document**. The document is now split into two documents.

Append two documents

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

1. Select **View** > **Show Selected Batch** > **All Documents** from the main menu, or click **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**. The document now appears in the list as a multi-page document. Web Verifier keeps the document name of the previous document for the new multi-page 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.
	Moves image to the left.
	Moves image to the right.
	Moves image upwards.
	Moves image downwards.

Document View - document area

This area shows the first page of the document selected in the batch structure.

About the Verification View classification window

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.

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

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 + 0	Print
CTRL + ALT + HOME	First Document
CTRL + Alt + PAGE DOWN	Previous Document
CTRL + Alt + PAGE UP	Next Document
CTRL + Alt + END	Last Document
CTRL + 8	Append Document
CTRL + 9	Cut Document
CTRL + ENTER	Accept/Reject next unsure page
CTRL + SPACE	Select next unsure page
CTRL + 1	Show Batches
Shift + ADD	Zoom In

Keyboard Shortcut	Command
Shift + 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 + H	Fit to Height
CTRL + W	Fit to Width
CTRL + F	Best Fit
CTRL + J	Increase image area
CTRL + K	Decrease image area
F7	Reclassify Document Manually
F9	Move to Exception State
CTRL + E	Release Exception Batches
CTRL + T	Correct Table
CTRL + Q	Switch Table Highlighting

Verification view classification window toolbar buttons

The toolbar provides quick access to some frequently used commands.

Button	Description
	Not available in classification view.
	Displays the Batch View.
	Not available in classification view.
	Displays 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.
	Highlights candidates – not available in classification view.
	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.
	Not available in classification view.

Button	Description
	Keeps the established zoom settings on each document you view in the batch.
	Moves image to the left.
	Moves image to the right.
	Moves image upwards.
	Moves image downwards.
	Displays the first page in document. This button is only enabled if the current document has more than one page. If you click on it, the first page is displayed.
	Displays the previous page in document. This button is only enabled if the current document has more than one page. If you click on it, the previous page is displayed.
	Displays the next page in document. This button is only enabled if the current document has more than one page. If you click on it, the next page is displayed.
	Displays the last page in document. This button is only enabled if the current document has more than one page. If you click on it, the last page is displayed.
	Rotates current page clockwise.
	Displays the first document in the batch.
	Displays the previous document in the batch.
	Displays the next document in the batch.
	Displays the last document in the batch.

Verification view classification window document area

This area shows the current document. All words that are recognized during the OCR process are highlighted by default.

About the class selection list

When filled, this box shows the classification result of the current document, or it is empty if there is no result determined. If you open the list, you will see all available classes.

Set or change a classification result

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

- Click on the arrow on the right side of the list box to open the list, and then select a class.

  Note: If you know the correct class name, you can type the first characters and wait until the system automatically displays the full class name.

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 displays automatically if the next document that is to be processed requires a correction of the extraction result. Whether this is the case depends on the state of the document.

Open the verification view indexing window

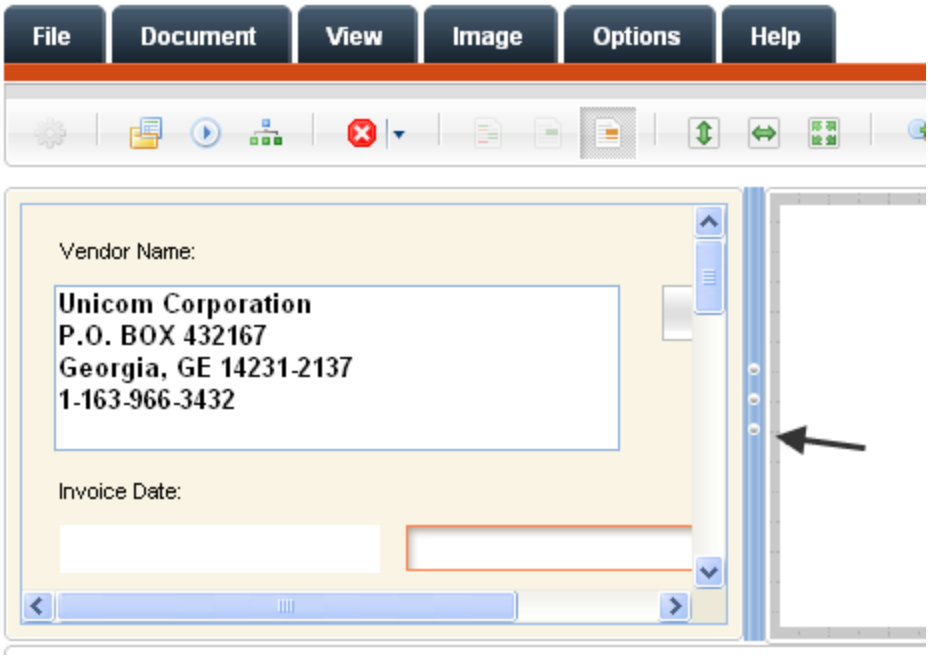
To display Verification View, complete the following steps.

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.



Indexing window keyboard shortcuts

The following keyboard shortcuts are available in the indexing window.

Keyboard Shortcut	Command
CTRL + 0	Print
CTRL + ALT + HOME	First Document
CTRL + ALT + PAGE DOWN	Previous Document
CTRL + ALT + PAGE UP	Next Document
CTRL + ALT + END	Last Document
CTRL + 8	Append Document
CTRL + 9	Cut Document

Keyboard Shortcut	Command
CTRL + ENTER	Accept/Reject next unsure page
CTRL + SPACE	Select next unsure page
CTRL + 1	Show Batches
SHIFT + SUBTRACT	Zoom Out
SHIFT + ADD	Zoom In
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 + H	Fit to Height
CTRL + W	Fit to Width
CTRL + F	Best Fit
CTRL + J	Increase image area
CTRL + K	Decrease image area
F7	Reclassify Document Manually
F9	Move to Exception State

Keyboard Shortcut	Command
CTRL + E	Release Exception Batches
CTRL + T	Correct Table
CTRL + Q	Switch Table Highlighting

Verification view indexing window toolbar buttons

The toolbar provides quick access to some frequently used commands.

Button	Description
	Not available in Indexing Mode.
	Displays the Batch View.
	Not available in Indexing Mode.
	Displays 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 supervisor to determine which exceptions to use. Note: 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.
	Highlights Candidates – Clicking this highlights all areas in yellow that were taken into account to fill the current field.
	Fits the current image to the height of the window.
	Fits the current image to the width of the window.

Button	Description
	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.
	Keeps the established zoom settings on each document you view in the batch.
	Moves image to left.
	Moves image to right.
	Moves image upwards.
	Moves image downwards.
	Displays the first page in document.
	Displays the previous page in document.
	Displays the next page in document.
	Displays the last page in document.
	Rotates current page 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.

Button	Description
	Displays the last document in the batch, and the application switches to Browsing Mode.
	Apply global extraction.
	Copy to global knowledge base.
	Correct tables.

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, check box, list box, or Yes/No field. A form may also contain buttons. A form can contain any of the following elements.

Toggle Read only button

Enables or disables the read-only mode for all fields. Fields that are designed as read only for the project are not affected by this button.

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 check boxes and combo boxes.

Check boxes

These toggle selections of data input, such as On/Off or Yes/No. Check boxes 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

They fire actions for a new script event.

Tables

Relevant when table extraction is configured. The Web 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 markers are used to indicate the nature of the field.

- The currently selected field is indicated by a red frame.
- Valid extracted fields are framed blue.
- Fields that need to be validated because they were extracted with low confidence are marked by a red triangle in the right upper corner.

Table fields provide additional highlighting options.

- Click on a table cell to highlight the appropriate document area.
- Click on a column header to highlight the appropriate document area.

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 current input area

Web Verifier supports the direct view snippet, which allows you to see an increased area of the current field to verify.

This direct view control, similar to the Verifier Thick Client, is positioned at the bottom of the extraction verification screen and allows you to see an increased image area selection of the current field candidate.

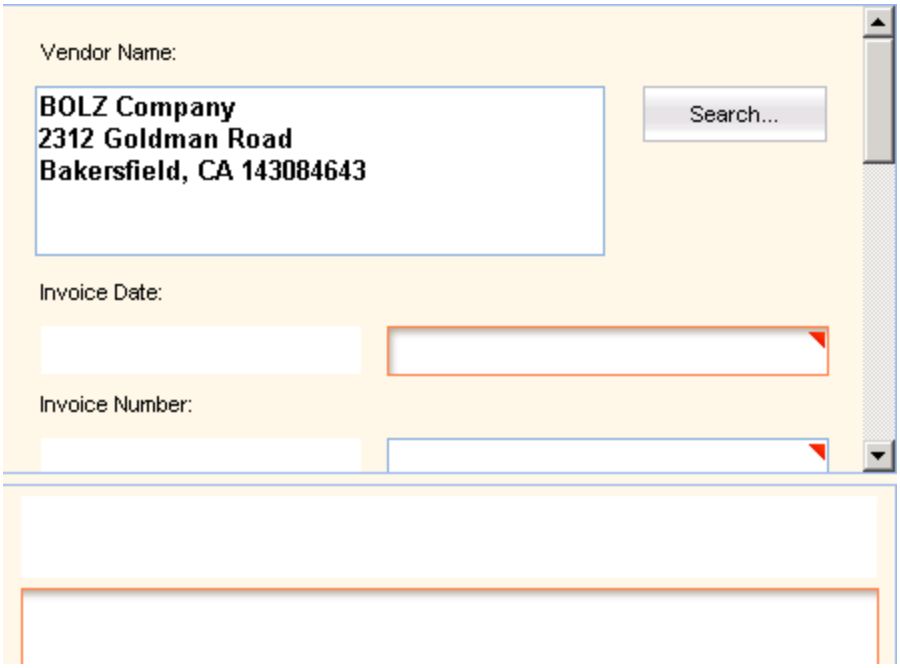
The current input area shows enlarged information for the currently selected field and provides a spacious edit box. A snippet also shows an enlargement of the document area that was used to fill the field. For the extracted data, color coding is used the same way as in the field area, and you can edit the data here.

Verification view indexing window document area

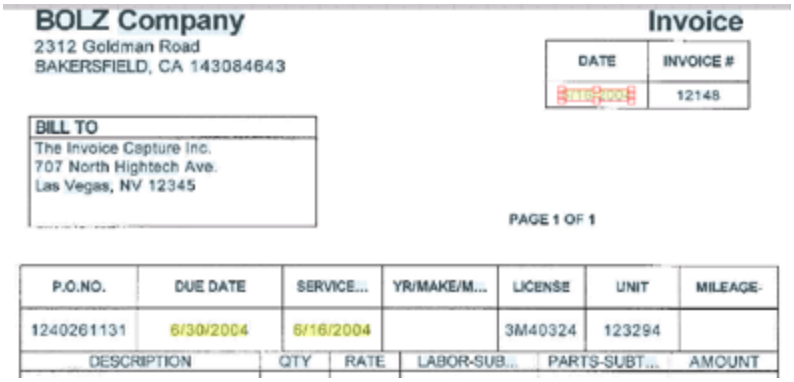
The document area shows the currently selected document or page along with highlights.

Note: The highlighting colors may vary due to customization.

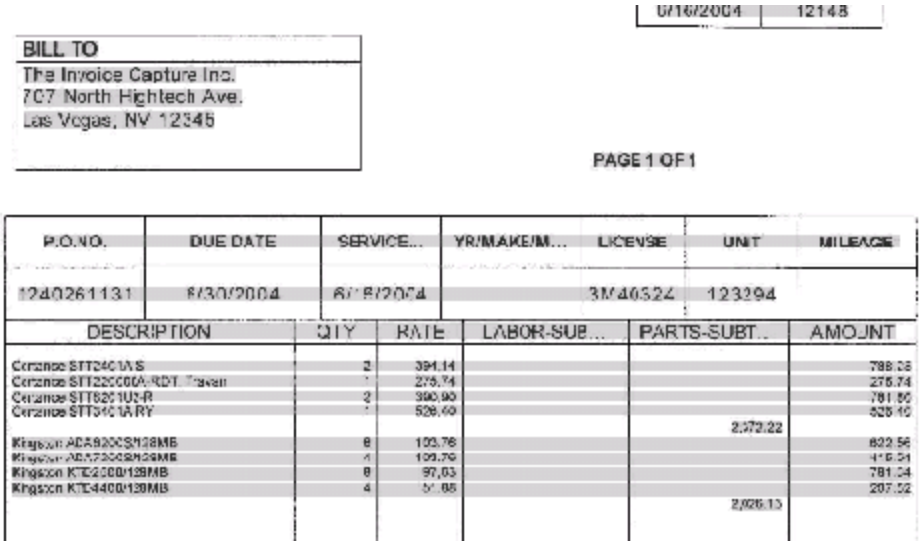
Red framed areas indicate the selection for the currently selected field.



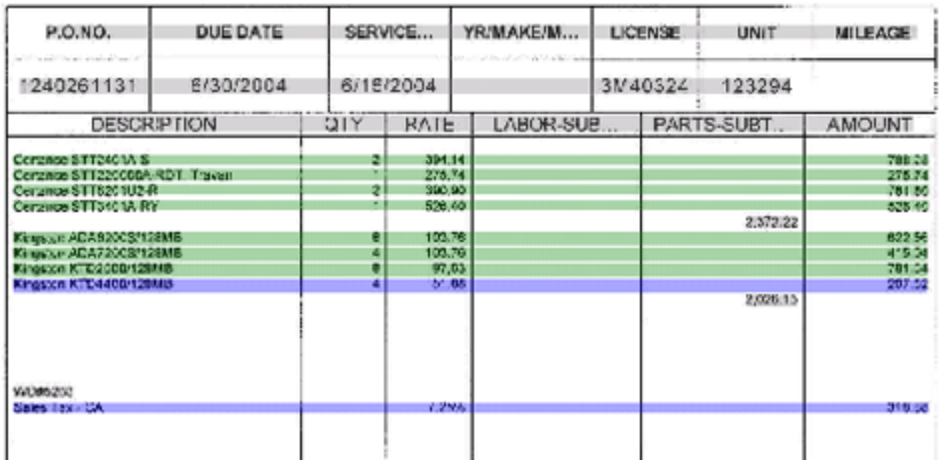
Yellow areas were considered as candidates, but another candidate seemed more likely. If the extraction result is invalid or wrong, these areas may point to the correct indexing data.



If you launch the table correction mode, the unlearned lines are shown in grey.



As soon as the lines are learned, the highlighting of the correctly recognized lines will turn to green, or blue if the line is selected with low confidence.



Print a document

To print a document, complete the following step.

- Select **File > Print**.

  Note: Web Verifier allows all the pages of the currently selected document to be printed by selecting the **All** button, or a desired page or page range to be specified in the **Pages** field of the **Print** dialog.

  Note: This function is available in all modes of Web 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 > Print Setup**.

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

When you have configured your print settings, Web 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, Web 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	Allows you to select the desired range of document file pages to be printed in addition to the form content, if enabled in user settings.
Print form	Activates printing of the verification form, which is turned on by default. The following options below are enabled only if the present one is activated.
Print hidden fields	When selected, Web Verifier prints not only the fields visible on the current verification form, but all the fields available in the loaded document.
Print table fields	When deselected, Web 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 “Print table fields” option is turned on. When turned on (default), Web Verifier prints column header on each page. This option is useful for printing of long tables.
Always show this dialog when printing	If this check box is checked, the Print Setup dialog will show when users press Print.

Work with Web Verifier

About the classification and indexing results

After classification, the result may be one of the following options.

- Documents that have not been classified display in the classification window. After 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 display in the indexing window.

Classify an incorrectly classified document

Documents that were incorrectly classified display in the indexing window. To classify a document, complete the following steps.

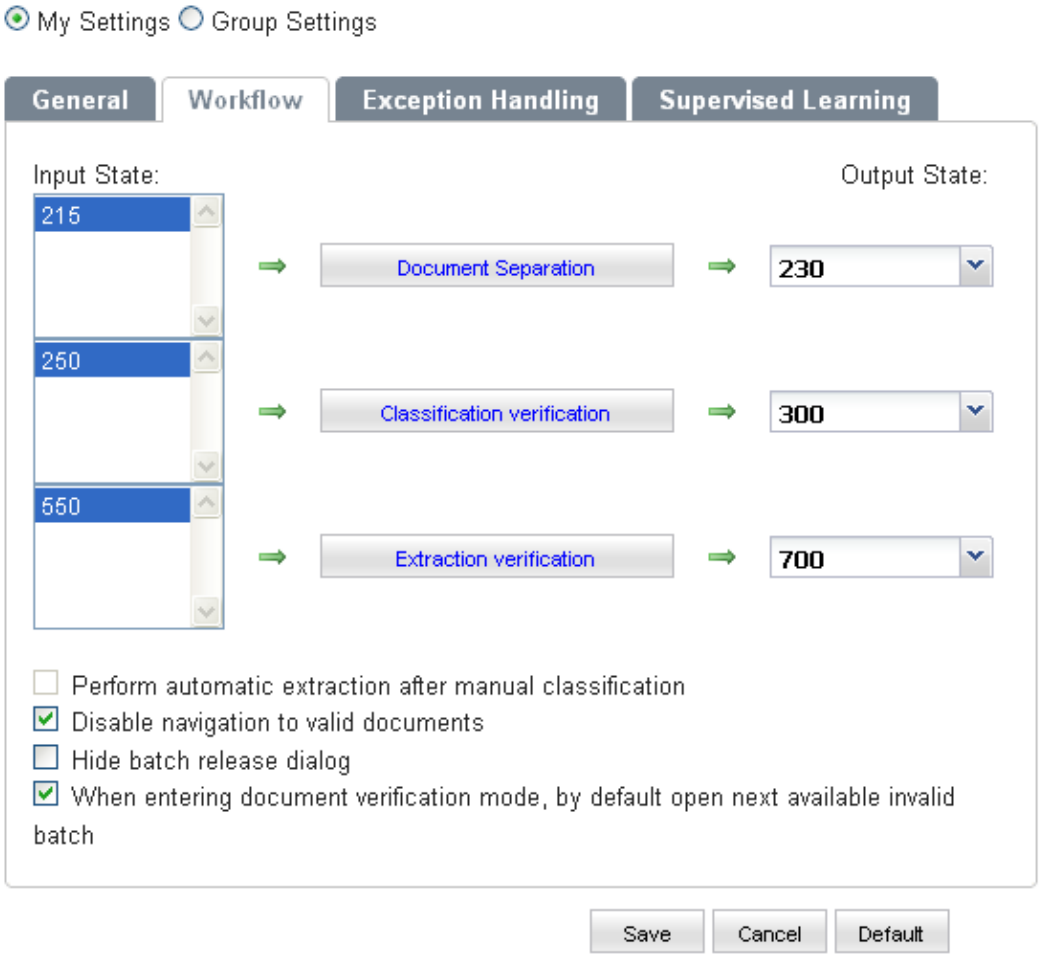
1. Select **Classify Document Manually** from the Options menu to open the classification window.
2. Correct the class and confirm by pressing `ENTER`. This displays the indexing window.

   Note: 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 manually index the document.

Page separation workflow in Web Verifier

Configure settings for document separation

Define the input state for Document Separation in the “Workflow” tab. The final step is to activate this workflow step by clicking Document Separation.



About manual correction of automatic document separation

If during automatic document separation in Runtime Server there was at least one unsure page-level decision for a batch of documents, the whole batch returns the “failed document separation” state.

Such a batch is supposed to be manually reviewed and, if required, corrected in the Web Verifier and Verifier applications. It is labeled by the icon.

The automatic document separation results can be corrected in Document Browsing mode of Web Verifier. When the next batch is opened, the system automatically displays the first uncertainly split/merged page.

Here, you have the following options.

- Toggle the unsure status
- Split the document into two separate documents
- Merge selected document with the previous one
- Go to the next unsure page

About toggling the unsure status

You can toggle the unsure status of a page. This sets the page to the **manually accepted** state or to the **manually rejected** state respectively.

There are three different states of page correction status.

- The blue page icon for extracted with high confidence by the engine.
- The blue page icon with a red question mark for extracted with low confidence by the engine (unsure).
- The blue page icon with green check sign for manually accepted/corrected by the Verifier user.

These states remain the same after you close the batch in Web Verifier and then other users can review them. If all pages of a document become accepted (the pages extracted with high confidence are accepted by default), the document is redirected to successful document separation state. Otherwise, if at least one of the pages becomes manually rejected, the whole document gets the lowest **page separation failed** state configured in Web Verifier settings.

Split the document into two separate documents

You can split the document into two separate documents. The “top document” receives all the pages above the currently selected one while the “bottom document” receives all the pages below, including the currently selected page. In this case, the currently selected page and the preceding page automatically get “manually accepted” page correction status. To split the document, complete the following step.

- From the **Document** menu, select **Cut Document**.

Merge the selected document with the previous one

You can merge the selected document with the previous one. In this case, the first page of the currently selected document and the last page of the preceding one automatically get “manually accepted” page correction status. To merge the selected document, complete the following step.

- From the **Document** menu, select **Append Document.**

Go to the next unsure page

You can go to the next unsure page. This action selects the next unsure page to verify (the one with a red question mark) without changing any page states. To advance to the next unsure page, complete the following step.

- From the Document menu, choose **Select Next Unsure Page**.

About page view modes

When verifying the correctness of automatic page separation in Web Verifier, you can switch to a more convenient page view mode. For example, you can have two consecutive pages displayed simultaneously.

There are four different page view modes available in Web Verifier.

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

In multi-page view, page presentation behavior is as follows:

- In the two-page mode, the selected page is on the right (or bottom in 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.

Change view modes

To change view modes, complete the following step.

- From the **View** menu, select **Multi-Page View**, and then select the option you want.

Reorder pages

You can reorder the pages within the document tree. To move a page to another position, complete the following steps. Note that after reordering a page, the document has to be reprocessed.

1. Perform a long left-click on the desired page.
   A tooltip appears indicating to specify the new position.
2. Drag the page to the new position.
   A dynamic green position line indicates the new position.
3. Click **Yes** on the reprocessing notification to complete the page reordering.

About manually correcting classification results

You can manually correct classification results if your Web 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 Web Verifier Settings dialog box.

Note: If you do this task regularly, you may want to apply the appropriate filter in 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, choose one of the following options.
   - To open the list and select a class, click on the arrow on the right side of the list box.
   - To browse the list of classes and make your selection, use the arrow keys. 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 displays automatically. Proceed as described above.
5. When the verification for all documents in the batch is complete, the application prompts you to select what you want to do next. Choose one of the following options.
   - To release the current batch and open the next batch that needs verification, select **Verify the next invalid batch on the list.**
   - To release the current batch and display the Batch View where you can select the next batch, select **Close batch and return to the batch list.**
   - To change verification forms, select **Verify this batch with the next verification form.**
6. Confirm your choice by clicking **Yes** or **No**.

About processing documents classified to obsolete document classes

Web 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 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 manually correcting extraction results

Manual correction of extraction results is available if the Web Verifier workstation is configured with the following settings.

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

To determine your settings, check the workflow tab of the Web Verifier Settings 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, extraction only**.

Correct invalid results

To correct invalid results, complete the following step.

- In Batch View, check the state column to find a batch you can verify. Clicking on it immediately starts the verification.

  The Verification View opens in Verify mode, and the first invalid document displays. The application places the cursor in the first invalid field.

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 check boxes or combo boxes. The field types for validation include Read Only, Multi-line, Combo Box, and Check Box. You can also add a button to a form to fire actions.

Elements of a form can include:

Form fields

Form fields display extracted data. You can also enter and edit data during manual indexing. You can use form fields to create check boxes and combo boxes.

Labels

Labels identify form fields, viewers, and tables.

Viewers

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

Buttons

Buttons fire actions for a new script event.

Tables

Tables are extracted from documents.

The following is a list of field types and their descriptions:

Read Only

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

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 list that lists predefined strings related to the verification document. It can either show the nearest values automatically or show only selected values.

Check Box

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

About editing text fields

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

Options for character changes include multi-line fields, combo boxes, and check boxes. 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.

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

The Check Box provides an either/or option that toggles table data entry choices on and off. For example, a Yes/No check box where selecting Yes would bring up data entry related to the verification.

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

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 is only updated if the candidate appears once in the document; otherwise, when the word for the field is inserted by auto-completion, the viewer will be blank.

Auto-completion does not work on formatted text and characters that were 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. 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 copy it to the cell, append or replace by it a word in a cell. The Copy feature fills the cell with the text.

Words that are not candidates for a cell

If the word does not belong to cell areas, it will display 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 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.

Note: It is not possible to append words from other or remote columns. If you select words or figures from different columns, the command **Append Field Text by Word** will replace the selected field’s content.

Use a word that is a candidate for a field

If the word is a candidate for a field, you can copy it into the field box. A candidate is one that matches the learned selections for the field. To copy text to the field box, complete the following step.

Prerequisite Make sure that the word fits the format analysis rules defined for that field.

- Click on the text you want to copy. A box appears around the word. Double-click on the box or right-click in the box and select **Copy to Current Field**.

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

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 still use it to fill a field. For example, a field named “sales total” might be replaced by “invoice total.” To use text for filling a field, complete the following step.

Prerequisite Make sure that the word fits the format analysis rules defined for that field.

- Drag a box around the word you want to use or click on it. Double-click on the word in the box, or right-click in the document and select Align & Copy to Current Field.

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

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.

   After the validation is finished, the cursor automatically moves to the next invalid field, regardless of whether this field is in the same or next invalid document. If you leave the document this way, it is validated automatically. In the next field, proceed as described in the previous sections. When all documents in the batch are validated, the application prompts you to select what you want to do next.
2. The following choices display. Select the next step by clicking **Yes** or **No**, or click **Details**.
   - **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.
- To determine your settings, check the Workflow tab of the Settings dialog box.

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: 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 manual indexing

Manual indexing is done if the Runtime Server was not configured to do the extraction step. Your input consists of batches with valid classification results, but no fields have been filled so far.

For manual indexing, your Web Verifier workstation is configured as follows:

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

To determine your settings, check the workflow tab of the Web Verifier Settings 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, extraction only**.

Manually index a document

To index a document manually, complete the following steps.

1. In Batch View, check the State column to find a batch you can verify and then click on a batch to select it.
2. The Verification View opens in Verify mode and the first document displays. Web Verifier places the cursor is in the first field.
3. Enter the value for the first field.
4. Press ENTER to validate your entry. The cursor automatically moves to the next field, regardless of whether this field is in the same or next invalid document. If you leave the document this way, it is validated automatically. In the next field, proceed as described above.
5. When you have finished indexing all documents in the batch, the application prompts you to select what you want to do next. The following choices display.
   - **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 Batch View where you can select the next batch.
   - **Verify this batch with the next verification form**. This changes verification forms, using the next verification form.
6. Move to the next step by clicking **Yes** or **No**.

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 and then select a batch by clicking on it.

   The **Verification View** with the first document requiring correction displays.
2. To display the first document in the batch, press the  button.
3. You may encounter a document that has been classified incorrectly. To correct this result, from the **Options** menu, select **Reclassify Document Manually** to open the classification window. To correct the class, select the corresponding entry from the list at the bottom, and then confirm by pressing ENTER.

   This displays the indexing window again.
4. 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.
5. To get to the next document, press the  button.
6. Proceed as described above until you reach the last document.

About working with tables

In Intelligent Capture, it is possible that any given table was trained by an extraction engine, Brainware Table Extraction, or the traditional Table Analysis Engine.

A Verifier user does not necessarily know which one was used, and the user cannot see a difference in Web Verifier. The process of table fields is similar, regardless of whether you're using Verifier or Web Verifier.

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

About automatic training and extraction of verified table data

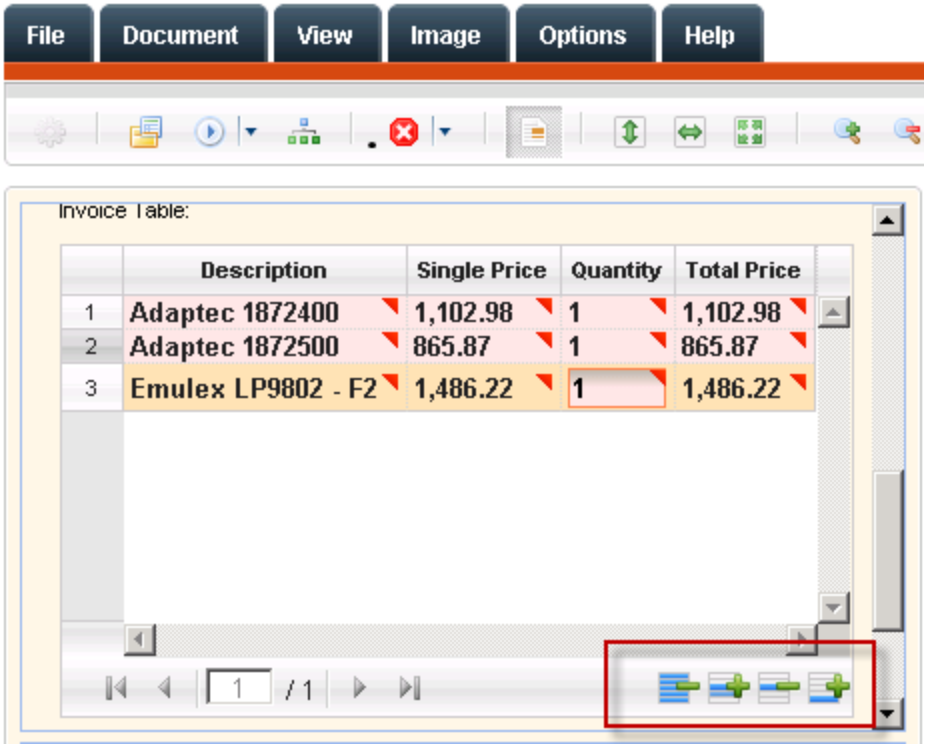
Brainware Table Extraction supports automatic learning of verified table data.

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

Traditional training and correction methods

Correcting table structure

You may also need to correct the table structure. Under the table within the verification mask, there are buttons for modifying the table structure.



The available commands are summarized in the following table.

Button	Command	Description
	Delete all rows	Deletes all rows from the table. You will need to relearn the lines on the invoice.
	Insert new row above	Inserts an empty row above the current one.
	Delete selected row	Deletes the currently selected row. To select a row, click on its number. The appropriate line on the invoice is unlearned and loses its color highlighting.
	Append	Appends an empty row at the bottom of the table.

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: This functionality is available for Supervised Learning verifiers.

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 Web Verifier Options menu or with the context menu options Show Lines and Show Columns.

If you have documents with many line items, you will benefit from the table navigation feature. In the web.config file, you are able to define how many lines should be displayed on one table page. If this number is exceeded, you’ll be able to browse between the table pages by using the table controls below the table.

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 rule #1 for the lines classified to the first line type and mapping rule #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, do the entire required training first and then verify manually.

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

When working with interactive table extraction learn lines before you map 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, un-map the column and re-extract the table to re-map the column. Although it will seem as if you deleted the data, the data will actually still be there until you un-map 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 in order.

1. Show the first row sample.
2. Learn columns in the rows you learned.
3. Learn missing lines.
4. Learn and adjust the mapping of missing or wrong columns.
5. Manually correct table fields and validate a table.
6. Learn the document.

Show the first row sample

To show the first row sample, complete the following steps.

1. To select your Brainware Table Extraction table, click 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 when you double-click 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.

   Note: Turning off the **Correct Tables** button causes the loss of learned rows and columns. Turning the **Correct Tables** button back on leads to the loss of mapped table data.

4. Optional. To learn lines as primary lines, complete the following substeps.
   1. Right-click any line marked in gray in the Image Viewer.
   2. On the shortcut menu, select **Learn Line**.
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 in a single row.
   2. Right-click the selection.
   3. On the shortcut menu, select **Learn as Primary Line(s).**

   Note: All correctly selected primary lines are learned, and all other lines are similarly extracted and displayed. If some lines were not extracted, try relearning the lines alone or in a block, as defined in the following step.

6. Optional. To learn a block of lines as a table row, complete the following substeps.
   1. In the document viewer, draw a rectangular selection over the primary lines in a single row.
   2. Double-click or right-click on the selection.
   3. From the shortcut menu, select **Learn as Row**. All correctly selected primary lines are learned and all other lines are similarly extracted and displayed.

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

Note: Do not try to learn the rest of the missing secondary or primary lines now. Mapping is defined on the basis of line type. If you attempt to train all different line samples now, you would need to learn the columns mapping separately for every line class. If you first learn the column mapping for the row you just learned, the next time you learn another line sample, the engine tries to apply existing mapping rules for the newly learned row automatically, significantly reducing the time to train the table.

Learn columns in the rows you learned

Complete the following steps to highlight the row you learned.

1. Switch to the columns highlighting. Select **Switch Table Highlighting** from the **Option** menu or right-click a line and select **Show column** and mark the location of the first cell item in the row you learned.
2. Right-click the first cell item in a column. Choose **Map Column**, select the required data column, and click the selection.
3. Repeat this step for the rest of the cell items in the first row.

Learn missing lines

Complete the following steps to learn if text is missing from a row.

1. Switch back to the lines highlighting mode.
2. Mark the next missing row and learn it as 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.

Learn and adjust the mapping of missing or wrong columns

To learn and adjust the mapping of a missing or incorrectly mapped column, complete the following steps.

1. Return to columns mapping learning mode and look for incorrect 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.

Manually correct table fields and validate a table

To manually correct table fields and validate a table, complete the following steps.

1. Switch to the 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 reactivates extraction and replaces all your manual inputs instantly.
2. To validate the table, press ENTER.

   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. 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 secondary lines will be learned, and all other lines are similarly extracted and displayed. If some lines were not extracted, repeat the procedure described above.

Advanced learning: additional functions

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**.

   Web Verifier highlights the selected lines and lines similar to it in gray. Information from these lines is not extracted.

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 content of the current field. 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 change the Default Colors, Background Colors & 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

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.