Oracle® Fusion Middleware

Oracle WebCenter Forms Recognition Web Verifier User's Guide

12*c* Release 1 (12.2.1.4.0)

E93590-02

September 2019

Documentation for WebCenter Forms Recognition Web Verifier, that describes how to configure and use the application to verify documents, and to perform Supervised Learning Workflow.

ORACLE®

Oracle Fusion Middleware Oracle WebCenter Forms Recognition Web Verifier User's Guide, 12c Release 1 (12.2.1.4.0)

E93590-02

Contents

1 What Is Web Verifier?

Web Verifier is a web-based extension of the Verifier client application. 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.

1.1 Some Helpful Terms

This section provides some helpful terms when using Web Verifier.

Accumulated Learnset	A common learnset is known as Accumulated Learnset.
Automatic Supervised Learning	This process uses the Associated Search Engine to process, classify, and extract information.
Base class	The highest level of classification.
Batch	A batch is a stack of electronic documents. The electronic documents may be paper-based or files created using applications.
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. Also called a subclass.
Class	A set of documents that are grouped by common content. Each class usually has a mnemonic name that describes its contents.
Common Learnset	An accumulation of Local Learnsets.
DocClass	A parent document class.
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 WebCenter Forms Recognition documents consist of one or several pages, though the concept of a page is not really required for digital documents.
Export	The system no longer manages documents released by the export function.
Folder	In a business environment, folders are normally used to keep several documents together. WebCenter Forms Recognition does the same thing with folders. However, in the context of WebCenter Forms Recognition, 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.
Global Learnset	A general learnset that encompasses similar classes or projects. See also Local Learnset.
Learning	Given a view with a set of documents in vector representation and their class assignment, a neural network is created, so that the defined classes can be reproduced without error. This neural network is used in all subsequent classification tasks.
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.
Local Learnset	A 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.
Optical Character Recognition (OCR)	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 class, called children.
Project	A collection of customer settings for applications. Projects are created in Verifier and sent 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.
Workdoc	An internal structure representing the logical structure of a document. The Workdoc represents the data created during processing of a single document and contains all OCR and analysis results.
Analysis	Analysis means the document content is analyzed and a set of possible values for a field are generated called candidates

Classification	Classification means assigning one or more classes and corresponding confidence values to one or more unknown documents.
Evaluation	Evaluation means determining a class or the contents of a field from confidence levels, weights, or distances for classes or candidates.
Extraction	Extraction means automatic document indexing and its information that helps to sort, file, and search documents.
Importing	Importing means bringing documents into the system for management and processing.
Indexing	Indexing means assigning attributes to a document. This can either be done manually, semi-automatically (Smart Indexing), or entirely automatically (Extraction).
Validation	Validation involves confirming whether a processing result is correct.
Verification	Verification is related to quality assurance. It involves taking a processed document, checking the processing results, and correcting any errors.

1.2 Quality Assurance with WebCenter Forms Recognition

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

- Batches are the basic entity.
- WebCenter Forms Recognition 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 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*.

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 it could not be filled, its content does not comply with validation rules that have been defined, or its 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.

- WebCenter Forms Recognition 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.

1.3 About WebCenter Forms Recognition Workflow

In WebCenter Forms Recognition, the flow of incoming documents follows a sequence of standard processing steps. One of the objective is to get documents to their recipients as quickly as possible.

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 step.

- Verification of the extraction step.

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 user manually analyzes and resolves the problem. Invalid batches are presented to you in a task list, called the *Batch View*. The Verifier user will have to resolve each problem and validate each correction before the batch can be released. Subsequent automatic steps can be carried out only after release. Finally, when WebCenter Forms Recognition has finished processing a batch, the documents are sent to their recipients.

2 About Installation

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 WebCenter Forms Recognition context using the Designer application, and will assign at least one role to you.

3 Start Web Verifier

3.1 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.

3.2 About Logging into Web Verifier

Your user name and password are assigned to you by your project administrator.

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.

3.2.1 Login to Web Verifier

To log into Web Verifier, 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.

3.2.2 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* page again.
- You can switch to another user by opening the **File** menu and clicking **Re-login**. The standard Web Verifier *Login* page 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 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.
  ```

3.2.2.1 Login Using Windows Authentication

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

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

2. 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.

3. Complete your Windows credentials if the Windows authentication form appears.

4. When logged in with Windows authentication, the Web Verifier window shows the domain and computer name in the top-right corner.

3.2.3 Issues You May Encounter During Login

3.2.3.1 Maximum number of users exceeded

There is a maximum number of users allowed to access the Intelligent Capture database at one time, which depends on the licenses.

3.2.3.2 Session timeout

To maintain the workflow with several users, a Web Verifier session ends automatically after 20 minutes of no activity .

3.3 Exit Web Verifier

To exit Web Verifier, complete the following steps:

Click the **Logout** link in the top-right corner.

Close your web browser.

4 About Roles

Depending on their assigned role, Web Verifier users are able to complete the following tasks:

- Define, modify, and maintain the learnset.
- Collect and manage local training data.
- Propose learn set candidates to improve the performance.
- Verify the documents that Runtime Server could not automatically process.
- Access and change the batch filtering properties.
- Access and change the settings.

Note: **See** the *Users, groups, and roles* in the *WebCenter Forms Recognition Designer User Guide* for more information.

4.1 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.

- Select **Change Password** from the **Options** menu.

- Type in your existing password in the **Old Password** field.

- Enter and confirm a new password in the respective fields.

- Click **OK**.

5 About Configuring Web Verifier

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

- The 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, i.e. 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.

5.1 Open the Settings Page

The *Settings* page is where you configure Web Verifier. To open the *Settings* page, Click the **Settings** button in the toolbar, or go to the **Options** menu and select **Settings**.

5.2 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, they are assigned to one of the available groups with specific roles assigned. Depending on the roles the current user has been granted, different *Settings* views will be displayed.

Users who have been granted the *Administrator* role can assign settings to other user groups by selecting the **Group Settings** option, then selecting the group to configure from the dropdown list. The defined settings will then become the default settings for all users who are members of that group.

The **Supervised Learning** tab is only available for users who have been granted the *Administrator*, *Supervised Learning Manager* or *Supervised Learning Verifier* role.

5.2.1 About Specifying the Project File

The **Use Project File** option is for selecting the file name of the WebCenter Forms Recognition project used to process the documents. This contains the design of the indexing form to verify the extraction. The following scenarios are possible for the project loading procedure:

- When you log in to Verifier for the first time, the system prompts you to let you know that the project is loading.

- When you select a new project and click **Save**, the project loads after returning to the Batch View.

- When 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 opens for a different project.

5.2.2 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 or Traditional).
- Danish.
- Dutch.
- English.
- French.
- Finnish.
- German.
- Italian.
- Japanese

- Korean.
- Norwegian.
- Polish.
- Portuguese.
- Romanian.
- Russian.
- Spanish.
- Swedish.
- Turkish.

5.2.3   About 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 checkbox. This leaves you the option to refresh the *Batch View* manually, by clicking the **Refresh** button at the bottom of the batch list.

5.2.4   About Specifying Tabbing Behavior

Selecting the **Tab through invalid fields only** option allows the user to skip all the valid fields and focus on the invalid fields. To tab through invalid documents, press [Tab].

When the user presses the [Tab] key inside a table control, the system tabs through invalid table cells only. This option also applies for tabbing in a backwards direction when the user presses [Shift] + [Tab].

5.2.5   About Enabling Auto-Completion

You can select the **Enable Text Typing Auto-Completion** option to speed up typing. When you start to type, auto-completion completes the word, suggesting the best match among all of the words or candidates available after OCR and Format Analysis.

You can also enable this function for **Use Currently Opened Page Only**. This checkbox is only available when the auto-completion option is checked.

5.3   About Group Settings

Group Settings is available for users who have the *Administrator* role assigned. Other users configure their own settings under **My Settings**.

In the View tab, the administrator can configure settings of different groups here by choosing the specific group from the dropdown list.

5.4   About the Workflow Settings

The settings allow to define the workflow of documents that are processed by Verifiers. After each processing step, output states are assigned to batches, which distinguish success from failure.

5.4.1 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:

1. 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.

2. 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.

3. To set an output value, select it from the **Output** box.

4. To force WebCenter Forms Recognition 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.

5. To skip validated documents and show the next invalid document when you click the **Next Document** button, select **Disable Navigation to Valid Documents**.

6. To go directly to the next invalid batch without the confirmation release dialog, select **Hide Batch Release Dialog**.

7. To open the first available invalid batch in the list, select **by Default Open the first available invalid batch and not the selected one**.

Note:	When this option is cleared, the batch selected in the list opens.

5.5 About Exception Handling Settings

The Exception Handling tab display settings that will help verify a document that is not processing normally.

A document with an error or missing information may not be suitable for verification. Moving the document into an exception state will flag the batch for issues.

Having a mechanism for handling unexpected errors allows operators to remove the batch from their task list. Then operators can manually assign special states to documents.

5.5.1 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:

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

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

To set the description label, right-click on the existing label and select **New Description** from the popup menu.

Type the label into the corresponding field and confirm.

> **Note:** The maximum length allowed for a description is 128 characters.

5.5.2 Configuring Exception Handling

The following settings are available for exception handling:

Before Moving a Document to an Exception State, Save It Automatically	Saves a document automatically before moving it to an exception state. This applies only to the respective current document.
Create New Batches with Documents Marked for Exception Handling	When this option is selected, the documents that are marked for exception handling will be moved to an exception batch. ▪ A batch is created for each exception code. ▪ The new batch receives a new batch ID. ▪ Documents from all verified batches are moved to the same exception batch in the *Batch View*. ▪ These batches can be released manually or automatically. When this option is disabled, 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	When this option is selected, an exception batch is released once it contains more than the defined number of documents, or is older than the defined number of minutes. This allows critical exception documents to be processed without waiting for manual intervention. Exception batches will also be released when user has exited the application and logged in again.
By default, set exception mode to Batch	Selecting this option changes the scope of the command **Move to Exception State on Options** menu to **Batch automatically**.

Allow user selection of exception mode (Batch vs. Document)	This setting enables the dynamic changing of the exception mode on the Options menu, document verification view. By default, this option switches on. Use this option together with the option above to preserve a specific exception mode for the different user groups.

5.6 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. The following settings are available for Supervised Learning:

Activate Supervised Learning Workflow	Supervised Learning Workflow allows the user to copy documents from current job to knowledge base job.
Knowledge Base Job	The database job name that 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 rates.
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 the **Nominate for Learning** button on the *Document View* page.
Automatically if Unclassified or More than ... % of Fields Are Invalid	Documents will be copied automatically to the knowledge base if they meet the configured criteria.
	If the **Only if Activated by the User** option is not selected, and the document does not meet the criteria to be copied automatically, you can still copy a document to knowledge base by clicking the **Nominate for Learning** button on the *Document View* page.

5.7 About Batch Filtering

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

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

The saved filtering settings apply to the current *Batch View*, and the application saves them for future sessions.

5.7.1 Configure Batch Filter Conditions

To configure your filter settings, complete the following steps:

- Select **Filtering** from the **Options** menu**,** or click on the **Batch Filter** icon. Users with the *SET* role assigned can also access the filtering window from within the *Settings* page.

- In the filtering configuration, 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.

- Click **Clear Condition** to clear the filter condition settings.

- Optional. Within the **Group Settings** section, administrators can easily define filter conditions for several user groups:

    - Select a group from the dropdown list.

    - Click **Configure Filtering**.

    - Establish the filtering conditions and confirm by clicking **Apply**.

        Click **Save** to save your settings. Users assigned to this group will now be presented with a subset of batches matching the filter conditions.

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

---

6   Getting Familiar with the User Interface

6.1   About the Batch List

The batch list shows a list of batches, which is your work list, and displays after starting the application.

6.1.1   Show Batch List

To access the *Batch View*, click the **Show Batches** icon.

Note: If Web Verifier is not yet configured, the list of batches shows no data.

6.1.2   Batch View Shortcut Keys

The following shortcuts are available in the *Batch View*:

Keyboard Shortcut	Command
[Ctrl] + [0]	Print.

6.1.3   Batch View Toolbar Buttons

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

6.1.4   Batch View Icons

In the table of batches, a batch is represented by a single row. For each batch, an icon indicates its 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 by another user or application. Therefore it is unavailable..

	Batch contains documents with exception statuses. To work on the batch, release it when it is unavailable.

6.1.5 Batch View Columns

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

Batch ID	A number that can be used to uniquely identify the batch.
State	An integer between **0** and **999** that indicates the progress of batch processing. The state also indicates whether the batch is ready for verification.
Priority	An integer between **1** and **9** that indicates how urgent it is that a job be finished, where **1** is the highest priority and **9** is the lowest.
Name	An arbitrary name that is easier to read than the batch ID. Because the name is optional, it might be missing or set to a default value.
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.
Documents	The value in this column indicates the number of documents inside the batch.
Last User / Module	The computer name of the operator who has processed the batch before and the name of the application that most recently processed the batch.
Last Access	Displays the date when the batch was last processed.
External Group ID	Optional. The group ID which has been assigned to a batch relating to security. Batches can be assigned to user group via a unique ID. This column is not displayed by default.
External Batch ID	Optional. The name of the batch group. This can be used to represent any piece of information you would like to associate with a batch. For example, an external system ID or storage box ID. This column is not displayed by default.
Transaction ID	Optional. The transaction ID assigned to a batch. This allows the developer to synchronize a newly created batch of documents with another external system. It can be used to identify originators of batch of documents. This column is not displayed by default.

| Transaction Type | Optional. The transaction type assigned to a batch. This allows the developer to synchronize a newly created batch of documents with another external system. It can be used to identify the types of documents in batches or the source of the documents. |
| | This column is not displayed by default. |

6.1.6   Sort in Batch View

You can sort any column in *Batch View*. To sort any item, 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 values toggle between ascending and descending order, whether numeric or alphabetical.

6.1.7   Select a Batch in Batch View

To select a batch in *Batch View*, click the batch you want to select in the table of batches. Depending on the status of the batch, either the classification or the extraction *Verification View* opens.

In Verifier, use the following keyboard shortcut to navigate through the list.

Keyboard Shortcut	Command
[Home]	Navigate to the first batch.
[End]	Navigate to the last batch.
[Down]	Navigate to the next batch.
[Up]	Navigate to the previous batch.
[Page Down]	Navigate one page down.
[Page Up]	Navigate one page up.

6.2   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.

6.2.1   Open Document View

To open document view, click the **Show Selected Batch** icon.

6.2.2   Document View Keyboard Shortcuts

The following keyboard shortcuts are available in *Document View*:

Keyboard Shortcut	Command
[Ctrl] + [0]	Print.
[Ctrl] + [Shift] + [Home]	First document.
[Page Down]	Previous document.
[Page Up]	Next document.

[Ctrl] + [Shift] + [End]	Last document.
[Ctrl] + [8]	Append document.
[Ctrl] + [9]	Cut document.
[Ctrl] + [Enter]	Accept/reject next unsure age.
[Ctrl] + [Space]	Select next unsure page.
[Ctrl] + [1]	Show batches.
[Shift] + [+]	Zoom in.
[Shift] + [-]	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 the image.
[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.

6.2.3 Main Toolbar Buttons

Button	Description
	Display a property sheet where you can configure Verifier.
	Display a property dialog where you can configure the batch filtering condition.
	Displays the *Batch View*.
	Start the verification of the currently selected batch. Depending on the batch state, the batch is either displayed in the classification window or in the indexing window. A dropdown 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.
	Display the first two pages of the selected batch horizontally or the first two pages of the selected document.
3×	Display the first three pages of the selected batch horizontally or the first three pages of the selected document.

	Display the first two pages of the selected batch vertically or the first two pages of the selected document

6.2.4   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 or set to a default value.
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 that you have access to.

6.2.5   Navigate in the Document View

To navigate in the batch structure, choose from 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.

6.2.6   About Splitting and Appending Documents

In the document list, you can split multipage 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.

6.2.6.1   Split a Multipage Document

To split a multipage document, complete the following steps:

- Select **View** ➤ **Show Selected Batch**, or click **Show Selected Batch**.

In the document list, click on the desired multipage document. This document must have at least two pages.

Right-click on the page and then perform one of the following actions.

  a. Select **Cut Pages into a New Document**. The document is now split into two documents.

  b. Select **Cut pages into new document keeping cover page**. This option splits a single document into several smaller documents and corresponding TIF files. It also includes the cover page of the original document as the cover page for the new documents.

> **Note:** To mark a page as a cover page, right-click on the first page of the document and select **Mark as cover page**.

6.2.6.2 Append Two Documents

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

  ▪ Select **View** ➤ **Show Selected Batch** ➤ **All Documents** from the main menu, or click **Show Selected Batch** ➤ **All Documents**.

Select the document to append to the previous document.

Right-click the document.

Select **Append This Document to Previous One**. The document now appears in the list as a multipage document. Web Verifier keeps the document name of the previous document for the new multipage document.

> **Note:** If the append operation is unavailable, that is because changes were made to the sorting list. Switch back to the original sequence.

6.2.7 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 downwards.

6.3 Print a Document

To print a document, from the **File** menu, select **Print** from.

Note: This function is available in all modes of Web Verifier with the exception of batch browsing mode.

6.3.1 Printing Verified Data Content

The amount of data on a printed form can be configured from the *Print Setup* dialog. To access the *Print Setup* dialog, select **Page Setup…** from the **File** menu.

The order in which the fields are printed is defined by the order of the fields configured in the project.

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:

Print Image	When selected, also prints pages of the document file.
Print Form	Activates printing of the verification form. The following options below are enabled only if the present one is activated.
	This option is selected by default.
Print Hidden Fields	When selected, Verifier prints not only the fields visible on the current verification form, but all the fields available in the loaded document.
Print Table Fields	When deselected, Verifier does not print table fields. Disabling this option might be useful for quick printing of documents with long tables.
Print column header on each printed table page	Enabled only if the **Print Table Fields** option is selected. When this option is selected, Verifier prints column headers on each page. This option is useful for printing long tables.
	This option is selected by default.
Always show this dialog when printing	If this option is selected, the *Print Setup* dialog will show when users print.

6.4 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.

6.5 Open the Verification View Classification Window

To display *Verification View*, complete the following steps:

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

6.6 Classification Window Keyboard Shortcuts

The following keyboard shortcuts are available in the classification window:

Keyboard Shortcut	Command
[Ctrl] + [0]	Print.
[Ctrl] + [Shift] + [Home]	First document.
[Page Down]	Previous document.
[Page Up]	Next document.
[Ctrl] + [Shift] + [End]	Last document.
[Ctrl] + [8]	Append document.
[Ctrl] + [9]	Cut document.
[Ctrl] + [Enter]	Accept/reject next unsure age.
[Ctrl] + [Space]	Select next unsure page.
[Ctrl] + [1]	Display the *Batch View*.
[Shift] + [+]	Zoom in.
[Shift] + [-]	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 the image.
[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 manually.
[F9]	Move to exception state.
[Ctrl] + [E]	Release exception batches.
[Ctrl] + [T]	Correct tables.
[Ctrl] + [Q]	Switch table highlighting.

6.6.1 Verification View Classification Window Toolbar Buttons

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. ■ 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.
	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 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.
	Rotates the 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.

6.6.2   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.

6.6.3 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.

6.6.4 Set or Change a Classification Result

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

6.7 About the Verification View Indexing Window

The indexing window displays fields and documents specific to your organization.

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.

6.7.1 Open the Verification View Indexing Window

To display *Verification View*, complete the following steps:

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

6.7.2 Increase or Decrease the Image Area

To increase or decrease the image area, drag the vertical split bar between the image area and field area either to the right or left.

6.7.3 Indexing Window Keyboard Shortcuts

The following keyboard shortcuts are available in the indexing window.

Keyboard Shortcut	Command
[Ctrl] + [0]	Print.
[Ctrl] + [Shift] + [Home]	First document.
[Page Down]	Previous document.
[Page Up]	Next document.
[Ctrl] + [Shift] + [End]	Last document.
[Ctrl] + [8]	Append document.
[Ctrl] + [9]	Cut document.
[Ctrl] + [Enter]	Accept/reject next unsure age.
[Ctrl] + [Space]	Select next unsure page.
[Ctrl] + [1]	Display the *Batch View*.
[Shift] + [+]	Zoom in.
[Shift] + [-]	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 the image.
[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 manually.
[F9]	Move to exception state.
[Ctrl] + [E]	Release exception batches.
[Ctrl] + [T]	Correct tables.
[Ctrl] + [Q]	Switch table highlighting.

6.7.4 Verification View Indexing Window Toolbar Buttons

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. ▪ 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.
	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 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.
	Rotates the 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.
	Apply global extraction.
	Add document to the knowledge based and nominate for learning.
	Correct tables.

6.7.5 Form Area

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

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

Form Fields	These are controls that are used to display and edit extracted data, and to enter data during manual indexing. You can use form fields to create checkboxes and combo boxes.
Checkboxes	These toggle selections of data input, such as on/off or yes/no. Checkboxes are derived from form fields. You can set up the caption with the text desired and select the default view.
List Boxes	These contain a selection list to use when verifying an item on the document. Used during manual verification, this selection works with automatic completion.

Labels	Labels are captions that help users to identify form fields, as well as viewers and tables.
Viewer	A viewer contains snippets of document areas, normally those that were extracted to fill fields or tables.
Buttons	Buttons fire actions for a new script event.
Tables	Relevant when table extraction is configured. The Verifier form supports multiple tables. However, even if you defined multiple tables, you can only display the first table on the verification form. You can display different tables on different forms.

6.7.6   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.

The following list explains the field types:

- A user cannot edit or select a **Read Only** field.

- An **Auto-completion** field enabled the user to edit text by typing the first few letters of a word until best matching candidate appears.

- A **Multi-line** field enables line wrap and displays a vertical scroll bar. This field type is a requirement for address analysis.

- A **List box** contains a selection list to verify an item in a document.

- **Checkboxes** are toggle selections for data input that derive from form fields.

6.7.6.1   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] keys 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.

6.7.7 About the Current Input Area

The input area displays information for the current field and provides an editing box, while displaying the document area that relates to the field.

Web Verifier and **Verifier** have a direct view control for viewing a large image area of the current field.

6.7.8 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 gray. 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).

7 Work with Web Verifier

7.1 About the Classification 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.

7.2 Classify an Incorrectly Classified Document

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

- Select **Classify Document Manually** from the **Options** menu to open the classification window.

Correct the class and confirm by pressing the [Enter] key. 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.

7.3 Page Separation Workflow in Web Verifier

7.3.1 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**.

7.4 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 should be manually reviewed and, if required, corrected in the Web Verifier.

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 or 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

7.4.1 Toggle 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 pages extracted with high confidence by the engine.

- The blue page icon with a red question mark for pages extracted with low confidence by the engine (unsure).

- The blue page icon with green check sign for manually accepted or corrected by the Verifier user.

7.4.2 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, select the **Cut Document** option from the **Document** menu.

7.4.3 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, select the **Append Document** option from the **Document** menu.

7.4.4 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, select the **Select Next Unsure Page** from the **Document** menu.

7.4.5 About Page View Modes

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

The following page view modes are available in Verifier:

- Single page view (default).

- Two pages displayed horizontally view.

- Three pages displayed horizontally view.

- Two pages displayed vertically view.

In multipage view, page presentation behaves according to the following modes:

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

- In the three-page mode, the selected page is in the middle, the first image shows the previous page of the current document or the last page of the previous document, and the third image shows the next page of the current document or the first page of the next document.

7.4.5.1   Change View Modes

To change view modes, select **Multi-Page View** from the **View** menu, and then select the option you want.

7.4.6   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.

Note: Reprocess the document after reordering the page.

7.5 Manually Correct Classification Results

To correct invalid classification results, complete the following steps:

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

- Open the selected batch in 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.

- To select a class, either:

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

  - Use the arrow keys to browse the list of classes and make your selection. The entries in the list are sorted alphabetically.

  - If you know the correct class name, type its first characters and wait until the system automatically displays the full class name.

- To confirm your selection, press the [Enter] key. The application validates the document and its state increases. The next document requiring verification is displayed automatically.

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:

- **Verify Next Invalid Batch in the List** releases the current batch and opens the next batch that needs verification.

- **Close Batch and Return to the Batch List** releases the current batch and displays the *Batch View*, where you can select the next batch.

- **Verify This Batch with the Next Verification Form** changes verification forms and displays the next verification form.

7.6 About Processing Documents Classified to Obsolete Document Classes

Deleting a class will make the class obsolete. Often with Supervised Learning workflow, classes become obsolete because the global project's configuration deletes or just does not insert the class.

The only way to process obsolete document classes is if the class still exists in the project that the document is processing with. Information saves internally about the former parent class assignment, which makes it possible to process obsolete document classes.

7.7 About Manual Correction of Extraction Results

To manually correct the extraction results, make sure to enable the extraction verification by checking the **Workflow** tab of the Web Verifier *Settings* or the *Verifier Properties* dialog box.

If you do this task regularly, you may want to apply the appropriate filter in the *Batch View*. From the **View** menu, select the **Batch Filter** option, then select **Batches to Verify, Extraction Only**.

7.7.1 Correct Invalid Results

To correct invalid results, complete the following steps:

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

- Open the batch in the *Verification View*.

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

7.7.1.1 Form Elements and Field Types

A form can include the following elements:

Form Fields	Display extracted data. You can also enter and edit data during manual indexing. You can use form fields to create checkboxes and combo boxes.
Labels	Identify form fields, viewers, and tables.
Viewers	Are sections of document areas, normally those that were extracted to fill fields or tables.
Buttons	Fire actions for a new script event.
Tables	Extracted from documents.

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

Read Only	When selected, information on a field is dimmed and cannot be selected or edited.
Auto-Completion	Enables you to edit text in a field by typing the first two letters of a word. Auto-completion finishes the word with the best matching candidates.
Multi-line Fields	Required in the context of address analysis but can also be useful in other cases. A multi-line field enables line wrap and displays a vertical scroll bar, if required.
List Box	A dropdown box that lists predefined strings related to the verification document. It can either show the nearest values automatically or show only selected values.
Checkbox	A toggle selection for one of two choices of the data input for a field, for example, yes/no.

7.7.1.2   About Editing Text Fields

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

You can use automatic character entry, when the auto-completion is selected in the form field Properties dialog box, to edit text fields and cells. Other options for character changes include multi-line fields, combo boxes, and 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 checkbox provides a binary option that toggles table data entry choices on and off. For example, with a **Yes** or **No** checkbox, checking **Yes** would bring up data entry related to the verification.

7.7.1.3   About Auto-Completion

Auto-completion helps to speed up typing. When you start to type, auto-completion 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 number. The auto-completion feature finds the best matched candidate suggested by the format analysis engine and places it in that field.

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

The auto-completion feature for a header field should 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.

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

7.7.1.4 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.

To use the Append feature, the selected word must appear on the same line as the existing cell text. If not, the selected word will replace the existing cell text.

- **Words that are candidates for cells:** If the word belongs in a cell area, you can append or replace a word in a cell. The Append feature takes the current word behind the candidate and appends the cell text. It places the text in the best location, either to the right or left of the word, and in the cell location based on the text or location of the word. The word belonging to a cell area displays in green when selected. Or, you can replace the text. In the search region, word candidates are all words that are not covered (by location) by other table cells and that have the same beginnings as the whole text of the cell
- **Words that are not candidates for a cell:** If the word does not belong to cell areas, it displays in orange when selected. Even if it is not a candidate, you can append or replace the word. Appending places the text in the best location, either to the right or left of the word, by text or location of the word. For example, a cell named "C2658" might be appended by "number" or you can replace the cell text and location by the text and location of a word.

7.7.1.5 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 steps:

1. Click on the text you want to copy. A box appears around the word.

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

2. 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.

7.7.1.6 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. To use text for filling a field, complete the following steps:

1. Drag a box around the word you want to use or click on it.
2. Double-click on the word in the box or right-click in the document and select **Append Field Text by Word**.
3. To replace text, select the word with mouse. A box will appear around the word. Double-click it, or select **Replace Field Text by Candidate.**

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

7.7.1.7 Insert Blocks of Text

Inserting blocks of text is helpful when working with multiple data verification elements. To insert blocks of text, complete the following steps:

- Click and drag over the desired text in the image viewer.

- Release the mouse button. A rectangle appears around the text.

- Adjust the rectangle, if necessary, by selecting the nodes at any corner.

- 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.

7.7.2 Finish the Validation

- After you correct a field, press the [Enter] key to validate it.

  The cursor moves automatically to the next invalid field regardless of whether this field is in the same invalid document, or in the next invalid document. Leaving the document this way validates it automatically.

- Click **Yes** or **No**, or click **Details**. Clicking **Details** reveals the following options:

  - **Verify Next Invalid Batch in the List** releases the current batch and opens the next batch that needs verification.

  - **Close Batch and Return to the Batch List** releases the current batch and displays the *Batch View* where you can select the next batch.

  - **Verify This Batch with the Next Verification Form** changes verification forms using the next verification form.

7.8 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.

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.

If you do this task regularly, you may want to apply the appropriate filter in the *Batch View* by selecting **Batch Filter** then **Batches to Verify** from the **View** menu.

7.9 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* page.

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**.

7.9.1 Manually Index a Document

To index a document manually, complete the following steps:

- In *Batch View*, check the **State** column to find a batch you can verify and then click on a batch to select it.

The *Verification View* opens in *Verify* mode and the first document displays. Web Verifier places the cursor is in the first field.

Enter the value for the first field.

Press the [Enter] key to validate your entry.

Note:	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.

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 in the List** releases the current batch and opens the next batch that needs verification.
- **Close Batch and Return to the Batch List** releases the current batch and displays *Batch View* where you can select the next batch.
- **Verify This Batch with the Next Verification Form** changes verification forms, using the next verification form.

Move to the next step by clicking **Yes** or **No**.

7.10 Check Entire Batches

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

- In *Batch View*, use the status value to determine a batch you can browse through.

Open a batch in *Verification View*. The first document that requires correction is automatically displayed.

- To display the first document in the batch, use the appropriate toolbar button.
- 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 box at the bottom, then confirm by pressing the [Enter] key.

- This displays the indexing window again.

- To correct extraction results, type your corrections into the corresponding field. If a field has been changed, its state is set to invalid. Press the [Enter] key to validate the field you modified.

- To get to the next document, use the appropriate toolbar button.

- Repeat the above steps as appropriate until you reach the last document.

8 About Working with Tables

You can verify table in the same way you would correct an invalid header field.

Additional methods to simplify manual table extraction, such as the **Correct Table**, are available.

8.1 About Automatic Training and Extraction of Verified Table Data

8.2 Traditional Training and Correction Methods

8.2.1 Insert Candidate Words in Table Cells

You can insert single words or append existing text in table cells. To insert words in table cells, complete the following steps:

- To select the text you want to insert in a table cell, complete one of the following actions:

    - Double-click on the word.

    - Right-click on the word in the image viewer and select the respective option.

    - If you have candidates, double-click the desired candidate to replace it.

To replace text with the new text, select **Copy to Current Field**.

To append text with the new text, select **Append Field Text by Word**.

8.2.2 Insert Non-Candidate Words in Table Cells

Even if the word does not belong to any candidates for the cell, you can insert single words, append or replace existing text in table cells. To insert words in table cells, complete the following steps:

- To append text with the new text, double-click the word or right-click in the image viewer.

- Select **Align & Copy to Current Field**.

- To replace text, select the word, and then select **Copy to Current Field** in the shortcut menu.

8.2.3 Correcting Table Structure

You may 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 table below:

Button	Command	Description
	Delete all rows	Deletes all rows from the table. You will need to relearn the lines on the document.
	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 document is unlearned and loses its color highlighting.
	Append	Appends an empty row at the bottom of the table.

8.3   Table Extraction and Correction

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

- About Learning Lines.
- Learning column mappings.

These are discussed in detail in the following sections.

Note that functionality is available for the Supervised Learning Verifiers.

8.3.1   About 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 the **Options** menu, or with the context menu options **Show Lines** and **Show Columns**.

8.3.2 About Learning Column Mappings

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 the first set of mapping rules for the lines classified to the first line type, and the second set of mapping rules 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.

8.3.3 About Correcting Fields in Tables Created with Brainware Table Extraction

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

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

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

Because of how 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 remap the column. Although it will seem as if you deleted the data, the data is still there until you un-map the column.

8.3.4 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:

- *Show the first row sample*.

- *Learn mapping in the row you learned*.

- *Learn missing lines*.

- *Learn and adjust the mapping of missing or wrong columns*.

- *Manually correct the table date and validate the table*.

8.3.4.1 Show the First Row Sample

1. Select your *Brainware Table Extraction* table by clicking any table field inside the table grid.
2. Click the **Correct Tables** button.
3. In the lines highlighting mode, use the **Learn as Row** function to show the row sample. This function automatically learns the first line as a *primary line* and the rest of the lines as

secondary lines. This function is also available by double-clicking on the selected row area. Select the whole first row and learn it.

Note: The visual indicators for valid, invalid and questionable table lines are as follows: valid lines are highlighted green, invalid lines are highlighted gray, and questionable lines are highlighted blue.

Deselecting the **Correct Tables** button causes the loss of learned rows and columns. Selecting the **Correct Tables** button again leads to the loss of mapped table data.

---

4. Optional. To learn a new line as a primary line, right-click on any line marked in gray in the Image Viewer and on the popup menu, select **Learn Line**.
5. Optional. To learn a block of lines as primary lines, draw a rectangular selection over the primary lines in the Document Viewer, right-click, and then select **Learn as Primary Line(s)**.
6. Optional. To learn a lines block as a table row, draw a rectangular selection over the required multi-line (or single-line) table row in the Document Viewer, double-click or right-click, and then select **Learn as Row**.

Note: Do not try to learn the rest of the missing secondary or primary lines now. 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.

8.3.4.2   Learn Mapping in the Row You Learned

1. Switch to the columns highlighting mode. 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.

8.3.4.3   Learn Missing Lines

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

1. Switch back to the lines highlighting mode.

Mark the next missing row and learn it as before.

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.

8.3.4.4   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:

▪ Return to column mapping learning mode and look for incorrect or missing mapping.

▪ Correct any missing mappings. 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.

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.

Repeat these steps until the data is completely extracted or cannot be learned correctly.

> **Note:** There is always a chance that you will not get 100 percent extraction results.

8.3.4.5 Manually Correct Table Fields and Validate a Table

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

- 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.

To validate the table, press the [Enter] key.

Optional. Click the **Add Current Document to Local Learnset** button to add the document to the Learnset and then learn it, which is known as cross-document learning. Do this if the system did not suggest learning the document automatically.

> **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.

8.3.5 Advanced Learning with Brainware Table Extraction

It is necessary to use secondary lines explicitly when a row begins on one page and ends on the next, or when learning a unmapped secondary line leads to an extraction. Refer to the following two cases to understand more about these circumstances.

- The table row begins on one page and ends on the next.

  If a table row begins on one page and ends on the next page, you must use the **Learn as Secondary Lines** function (in lines learning mode) to train missing secondary lines (on the next page).

  In this case, these secondary lines are placed right before the first primary line on the page. Mark all the secondary lines as before: right-click and select **Learn as Secondary Lines**.

  > **Note:** 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.

  .

- 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.

8.3.5.1   Learn a Block of Secondary Lines

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

- In the *Document Viewer*, use the mouse to draw a rectangular selection over the required secondary lines of a desired multi-line row.

Right-click on the selection.

On the popup menu, select **Learn as Secondary Line(s)**.

Note: Repeat the procedure if the lines did not extract.

8.3.5.2   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:

- Switch to *Lines Learning* mode and right-click on the line you want to unlearn.

- On the shortcut popup, click **Unlearn Line**. Unlearned lines change from green to gray.

8.3.5.3   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:

- Right-click on any learned line or draw a rectangular selection over the required lines.

- On the popup menu, select **Learn as Wrong Line**. The selected lines and similar lines to it are now highlighted in gray. Information from these lines will not be extracted.