

Oracle® Fusion Middleware

Oracle WebCenter Forms Recognition Designer User's Guide
12c (12.2.1.3.0)

E93591-01

July 2018

Documentation for WebCenter Forms Recognition Designer, that describes how to use the application to create, configure, modify and test projects.

E93591-01

Copyright © 2009, 2018, Oracle and/or its affiliates. All rights reserved.

This software and related documentation are provided under a license agreement containing restrictions on use and disclosure and are protected by intellectual property laws. Except as expressly permitted in your license agreement or allowed by law, you may not use, copy, reproduce, translate, broadcast, modify, license, transmit, distribute, exhibit, perform, publish, or display any part, in any form, or by any means. Reverse engineering, disassembly, or decompilation of this software, unless required by law for interoperability, is prohibited.

The information contained herein is subject to change without notice and is not warranted to be error-free. If you find any errors, please report them to us in writing.

If this is software or related documentation that is delivered to the U.S. Government or anyone licensing it on behalf of the U.S. Government, then the following notice is applicable:

U.S. GOVERNMENT END USERS: Oracle programs, including any operating system, integrated software, any programs installed on the hardware, and/or documentation, delivered to U.S. Government end users are "commercial computer software" pursuant to the applicable Federal Acquisition Regulation and agency-specific supplemental regulations. As such, use, duplication, disclosure, modification, and adaptation of the programs, including any operating system, integrated software, any programs installed on the hardware, and/or documentation, shall be subject to license terms and license restrictions applicable to the programs. No other rights are granted to the U.S. Government.

This software or hardware is developed for general use in a variety of information management applications. It is not developed or intended for use in any inherently dangerous applications, including applications that may create a risk of personal injury. If you use this software or hardware in dangerous applications, then you shall be responsible to take all appropriate fail-safe, backup, redundancy, and other measures to ensure its safe use. Oracle Corporation and its affiliates disclaim any liability for any damages caused by use of this software or hardware in dangerous applications.

Oracle and Java are registered trademarks of Oracle and/or its affiliates. Other names may be trademarks of their respective owners.

Intel and Intel Xeon are trademarks or registered trademarks of Intel Corporation. All SPARC trademarks are used under license and are trademarks or registered trademarks of SPARC International, Inc. AMD, Opteron, the AMD logo, and the AMD Opteron logo are trademarks or registered trademarks of Advanced Micro Devices. UNIX is a registered trademark of The Open Group.

This software or hardware and documentation may provide access to or information about content, products, and services from third parties. Oracle Corporation and its affiliates are not responsible for and expressly disclaim all warranties of any kind with respect to third-party content, products, and services unless otherwise set forth in an applicable agreement between you and Oracle. Oracle Corporation and its affiliates will not be responsible for any loss, costs, or damages incurred due to your access to or use of third-party content, products, or services, except as set forth in an applicable agreement between you and Oracle.

Contents

Contents.....	3
1 What Is WebCenter Forms Recognition Designer.....	7
2 About This Document.....	8
2.1 Intended Audience.....	8
2.2 Content Overview.....	8
3 Basic Concepts and Techniques.....	10
3.1 Start Designer.....	10
3.2 Roles.....	10
3.3 Login.....	11
3.4 Create Users and Groups.....	11
3.5 Windows Based User Authentication.....	15
4 Modes.....	17
4.1 What Are Modes?.....	17
4.2 Document Selection Mode.....	17
4.3 About the Definition Mode.....	22
4.4 Switch to Definition Mode.....	22
4.5 Modify the Processing Delay.....	22
4.6 Support of Non-Western Languages.....	23
4.7 Normal Train Mode.....	26
4.8 Verifier Train Mode.....	27
4.9 Runtime Mode.....	30
4.10 Verifier Design Mode.....	32
4.11 Verifier Test Mode.....	32
5 Working with Documents.....	34
6 Working with Projects.....	38
7 Classification.....	48
7.1 About Classification.....	48
7.2 Advanced Classification.....	48
7.3 Prepare Sample and Test Documents.....	48
7.4 Create a Project.....	49
7.5 Create Classes.....	49
7.6 Classification Methods.....	51
7.7 Select Classification Methods.....	51
7.8 Project-Level Standard Classification.....	52
7.9 Class-Level Sub-Tree Classification and How It Works.....	55
7.10 Stretching.....	56
7.11 Calibration.....	56
7.12 Classification Learnset.....	56
7.13 Brainware Layout Classification Engine.....	61
7.14 Layout Classify Engine.....	63
7.15 ASSA Classify Engine.....	63
7.16 Brainware Layout Classification Engine.....	64
7.17 Forms Classification.....	67

7.18 Phrase Classification.....	71
7.19 Language Classify Engine.....	76
7.20 Image Size Classification Engine.....	80
7.21 Template Classify Engine.....	81
7.22 Engine-Level Absolute Results.....	81
7.23 Classification Testing.....	82
7.24 Classification Problems.....	86
8 Recognition Techniques and Engines.....	88
8.1 FineReader OCR Engines.....	88
8.2 Kadmos 5 Engine.....	91
8.3 Cairo OMR Engine.....	93
8.4 QualitySoft Barcode Engine.....	95
8.5 Cleqs Barcode Engine.....	96
8.6 OCR Field-Level Digit Booster Engine.....	98
9 Data Extraction.....	101
9.1 About Threshold and Distance.....	101
9.2 Field Types.....	102
9.3 Text Fields.....	102
9.4 Table fields.....	102
9.5 Create a Field.....	102
9.6 Create a Table Field.....	102
9.7 Specify a Display Name for a Field.....	103
9.8 Delete a field.....	103
9.9 Check the Analyzed Table Content.....	103
9.10 Rename a Field.....	103
9.11 Specify the Analysis Engine.....	103
9.12 Project-Level Recognition Settings.....	104
9.13 Modify the Field-Level Recognition Settings.....	105
9.14 Specify OCR Settings.....	105
9.15 Despeckling.....	106
9.16 About Line Extraction.....	106
9.17 Associative Search Engine.....	107
9.18 About the Brainware Table Extraction Engine.....	113
9.19 About the Check Analysis Engine.....	119
9.20 About the Format Analysis Engine.....	122
9.21 Table Analysis Engine.....	128
9.22 Zone Analysis Engine.....	135
10 Validation.....	142
10.1 About Validation.....	142
10.2 About Standard Validation Engine.....	142
10.3 About Validation Settings.....	142
10.4 About Force Validation Feature.....	142
10.5 Derived Validation.....	143
10.6 About Validation Rules.....	143
10.7 About Document Validation.....	144
10.8 About Validation Rules for Classes.....	144
10.9 About Possible Limitations for Validating Fields.....	146

11 Evaluation.....	159
11.1 About Field Evaluation.....	159
11.2 Configure Brainware Extraction Engine.....	159
11.3 Brainware Field Extraction Engine for Generic Fields Extraction.....	160
11.4 About Training Header Fields in Normal Train Mode.....	162
11.5 Train Header Fields in Normal Train Mode.....	163
11.6 About Learning the Extraction.....	163
11.7 About Creating Learnsets.....	163
11.8 About Learning.....	166
11.9 Test the Extraction.....	166
11.10 Extraction Optimization.....	169
11.11 About Extraction Retaining.....	170
11.12 Text Field Candidate Evaluation.....	171
11.13 Table Field Candidate Evaluation.....	171
12 Setup Supervised Learning.....	173
12.1 Considerations and Constraints for Supervised Learning Workflow.....	173
12.2 Apply the Supervised Learning Workflow.....	173
12.3 Example for a Supervised Learning Workflow.....	174
12.4 Train the Base Class.....	176
12.5 Establish Multiple Classification Results using a Script.....	177
12.6 Training Other Fields.....	177
12.7 Create Derived Document Classes.....	178
12.8 Create Derived Document Classes with the Supervised Learning Workflow.....	178
12.9 About the Supervised Learning Workflow Process.....	178
12.10 Adding Supervised Learning to Brainware Layout Classification.....	179
12.11 Specify the Default Classification Result.....	179
12.12 ASE Pool Configuration.....	180
12.13 Options for Creating Derived Classes in Verifier.....	180
13 Verification.....	182
13.1 About Quality Assurance with Verifier.....	182
13.2 About Verification Projects.....	183
13.3 Verification Testing.....	195
14 Regular Expressions.....	206
14.1 What Are Regular Expressions?.....	206
14.2 Literal Characters in Regular Expressions.....	206
14.3 Operators in Regular Expressions.....	206
15 Templates.....	208
15.1 About Reusing Project Settings with Templates.....	208
15.2 Create an Analysis or Evaluation Template for a Field.....	208
15.3 Use an Existing Template.....	208
15.4 Display the Available Templates.....	209
15.5 Modify a Template.....	209
15.6 Rename a Template.....	209
15.7 Delete a Template.....	209
15.8 Export Templates to another Project.....	210
15.9 Import Templates from another Project.....	210

16 Document Export	211
16.1 About setting up the document export.....	211
17 Regional and Currency Settings	212

1 What Is WebCenter Forms Recognition Designer

Oracle WebCenter Forms Recognition allows you to process incoming documents automatically.

As part of the WebCenter Forms Recognition suite, Designer enables you to customize the automatic processing of paper and electronic documents. You can also test system settings prior to production.

You define how to classify documents are relevant in your enterprise, determine what data to extract, and how to verify the results. Designer stores the settings you define in project files. The project files automatically passed to WebCenter Forms Recognition Runtime Server for production processing.

In Designer, you can assign users to groups, and define what roles the members of those groups can perform when working with project files.

2 About This Document

2.1 Intended Audience

This documentation is for users who intend to create custom applications with WebCenter Forms Recognition Designer. It is expected that users will have experience and knowledge of Microsoft Windows operating systems and of Microsoft Windows applications.

2.2 Content Overview

This manual is organized into the following sections:

- [Section 3: Basic Concepts and Techniques](#) contains information about the user interface and basic concepts you need to understand to work with WebCenter Forms Recognition Designer. This section contains information about the setup of user accounts and how to export them to the Forms Recognition database. This section also contains a functional description of controls you need to work with Designer. You should be familiar with this information before you start working with more advanced options of WebCenter Forms Recognition.
- [Section 4: Modes](#) contains information on how to setup several modes of operation.
- [Section 5: Working with Documents](#) provides information on how to work with documents.
- [Section 6: Working with Projects](#) contains information on how to work with projects.
- [Section 7: Classification](#) contains information on how to work with different types of classification.
- [Section 8: Recognition Techniques and Engines](#) contains information about recognition techniques. You need this information to improve classification and extraction results.
- [Section 9: Data Extraction](#) contains information on how to set up data extraction in WebCenter Forms Recognition.
- [Section 10: Validation](#) provides information about confirming a data extraction is correct or incorrect.
- [Section 11: Evaluation](#) contains information about the evaluation mechanism in WebCenter Forms Recognition. You need this information to improve classification and extraction results.
- [Section 12: Setup Supervised Learning](#) provides information about Supervised Learning Workflow.
- [Section 13: Verification](#) contains information to check and correct the results of automatic document classification and data extraction. In addition, it enables you to manually index documents.
- [Section 14: Regular Expressions](#) contains information about regular expressions. You may need this information to improve extraction results obtained with format analysis.

- [Section 15: Templates](#) provides information about analysis and evaluation settings which can be reused elsewhere within or outside your project. You may need this information to save analysis and evaluation settings as templates.

3 Basic Concepts and Techniques

3.1 Start Designer

Launch WebCenter Forms Recognition Designer from the Windows Start menu as follows:

All Programs ▸ Oracle ▸ WebCenter Forms Recognition ▸ Designer

Note: You can run multiple instances of WebCenter Forms Recognition Designer simultaneously.

For more information on projects, refer to [Section 5: Working with Projects](#).

3.1.1 Create the Administrator Password

As an administrator, your first task is to change your password for each existing project. Otherwise, anyone can log in to your project and make changes that could potentially affect the entire workflow. To change your password, complete the following steps:

1. Load an existing project. Use **Administrator** (not case-sensitive) for user name, and leave the password blank.
2. From the **Options** menu, select **Users, Groups and Roles...**
3. On the *Project Authentication Properties* dialog, select the **Change Password** tab.
4. Select **Administrator** from the **User Name** drop-down list.
5. Enter and confirm the password in the appropriate fields.
6. Click the **OK** button.

3.2 Roles

There are six roles as described in the table below:

Role	Abbreviation	Description
Administrator	ADM	Used 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 groups.
Supervised Learning Manager	SLM	Used to define, modify, and maintain the learnset. This functionality is accessible only through Verifier.
Supervised Learning Verifier	SLV	Used to collect and manage local training data. Supervised Learning Verifiers are subject matter experts who can propose learnset candidates to improve system performance. This functionality is accessible only through Verifier.
Verifier	VER	Used to verify documents that could not be automatically processed. This functionality is accessible only through Verifier.
Verifier Settings	SET	Used to change the Forms Recognition Verifier configuration. This role

Role	Abbreviation	Description
		is given to users who are considered to have enough knowledge of the application to make changes that will be beneficial to all Verifier users.
Verifier Filtering	FLT	Allows for more flexibility and security when working with the Verifier client. The filtering tab in Verifier is unmapped from the settings dialog box, so that FLT role users are able to filter batches even if they do not have the SET role.

3.3 Login

This section provides information on how to log in to Designer and Verifier.

3.3.1 Login to Designer

To use Designer, you must be assigned to a group that has the *Administrator* role. User without this authorization is unable to use Designer.

As an administrator, you must log in to any Designer project you load.

You do not have to log in when you create a project, but you are essentially logged in as an administrator when you save the new project.

When you save a new project, you are logged in as the **Administrator** user. No password is required the first time you log in to the new project. Your first action should be to change the **Administrator** user's password.

3.3.2 Login to Verifier

Users must enter a username and password when loading a project in Verifier. Users who forget their passwords must contact their project administrator to have their password reset.

3.4 Create Users and Groups

This section provides information about how to create and manage users and groups.

3.4.1 Groups

- Users inherit the roles of their assigned groups.
- Groups can have one or more roles.
- A group must have at least one role.

3.4.1.1 Create Groups

By default, the first time the administrator logs in, the only existing user group is **Administrator**, and the only member of the group is the **Administrator** user.

To create new groups, complete the following steps:

1. Select the **Groups** tab.
2. Click the **New** button. The *Add New Group* dialog is displayed.
3. Enter a name for the new group in the **Name** field. The field is restricted to 20 alphanumeric characters.
4. For **Description**, select one of the four group roles options:

- Administrator
 - Supervised Learning Manager
 - Verifier
 - Supervised Learning Verifier
5. Click the **OK** button.

3.4.1.2 Delete a Group

To delete a group, complete the following steps:

1. On the **Options** menu, click **Users, Groups and Roles**.
2. On the **Project Authentication Properties** dialog box, on the **Groups** tab, select a group and then click **Delete**.

Note: You cannot delete the Admin group.

3. In the **Warning** dialog box, click **Yes**.

3.4.1.3 Assign Roles to Groups

When you first create a group, you can only associate one of the four profiles to it. However, you can assign or remove roles from an already-established group.

To add a role to a group, complete the following steps:

1. Select the role in the **Available Role** list and click **Add**.
2. Click the **OK** button.

After you have established user groups, you are ready to assign users to them.

3.4.1.4 Remove Roles from a Group

To remove a role, complete the following steps:

1. Select the role in the **Assigned Role** list and click **Remove**.
2. Click the **OK** button.

3.4.2 About Users

To open a project in Designer, you must log in with a user ID and, if specified, a password. A user must be a member of one or more groups. A group must have one or more assigned roles. You can create the following types of users.

- Project users that are only available in the project.
- Script imported users. For more information, refer to [Section 3.4.5: Import users and groups using the security update feature](#).
- Windows authenticated users, includes project users and script imported users. For more information, refer to [Section 3.5.3: Import Windows users in Designer](#).

3.4.3 Manage User Accounts

The following settings are user-specific:

- Local Project

- Common Folder
- Use database as knowledge source
- Select Job
- Assigned Groups

The following settings apply to all users:

- In the login dialog box, set the ability to remember passwords
- Allow Windows Authentication
- Allow Database Authentication

Note: From version 12c (12.2.1.3.0), all projects should have the **Allow Database Authentication** option enabled, and the users should exist in the database to login to the system.

3.4.3.1 Create New User

To create new user accounts, complete the following steps:

1. On the **User** tab, click **New**. The *Add New User* dialog is displayed.
2. Enter a username for the user and assign a default password in the **Name** and **Password** fields respectively. User names cannot exceed 17 alphanumeric characters and passwords cannot exceed 20 alphanumeric characters.
3. Click **OK**.
4. Optional. In the **Local Project** field, type or browse to a local project.
5. Optional. To specify the knowledge source location, complete one of the following sub steps:
 - a. In the **Common Folder** field, type or browse to a directory.
 - b. Select **Use database as knowledge source** and from the **Select Job** list, select a job.
6. From the **Available Groups** list, select a group and then click **Add**.

Note: Users who are not assigned to a group cannot log in to the project.

7. Optional. Repeat the previous step for each required role.
8. Optional. To remove an assigned group, from the **Assigned Groups** list, select a group and then click **Remove**.
9. Optional. To allow saving the password in Verifier, select **Enable Password Remembering for User Logon Dialog**.

Note: This setting has no effect if the database is in use.

10. Optional. Select **Allow Windows Authentication**.
11. Select **Allow Database Authentication**.
12. Click **Apply** and then click **OK**.

3.4.3.2 Delete a User

To delete a user, complete the following steps:

1. On the **Options** menu, click **Users, Groups and Roles**.
2. On the **Project Authentication Properties** dialog box, on the **Users** tab, select a user and then click **Delete**.

Note: You cannot delete the Administrator user.

3. In the **Warning** dialog box, click **Yes**.

3.4.3.3 Change a User's Password

To change a user's password, complete the following steps:

1. On the **Change Password** tab, check the **Decode Password** check box.
2. Enter and confirm the new password in the **New Password** and **Confirm New Password** text boxes.
3. Click **OK**.

Note: This only works for users that were created within the current project file. Users created through *Security Update* are not visible to the WebCenter Forms Recognition Designer.

3.4.3.4 Export Users and Groups

The definitions of all users' roles for a project can be exported in bulk to a binary file. This is done within the **Export** button on the **Users** tab of the *Project Authentication* dialog.

To export a list of users, complete the following steps:

1. On the **Users** tab, select the group to export and click **Export to File**.
2. Name the file and then click **OK**. By default, the file is saved as an authentication file with a **.sec** extension in the project root folder. You can change the target folder if necessary.

3.4.3.5 Import Users and Groups

The definitions of all users' roles for a project can be imported in bulk. This is done within the **Import** button on the **Users** tab of the *Project Authentication* dialog.

To import a list of users from another project, complete the following steps:

1. On the **Users** tab, click **Import from File**.
2. Select the appropriate file with **.sec** extension to import and click **Apply** and then click **OK**.

3.4.4 Export Users, Roles and Groups to the Forms Recognition Database

In Verifier, you have the ability to make the users, roles, and groups available in the WebCenter Forms Recognition database. This is necessary if a database platform is used by WebCenter Forms Recognition applications, like Web Verifier.

To make users, roles, and groups available in the database, use the **Export to Database** button in the project authentication dialog, which is available through the **Users, Groups and Roles...** option of the **Options** menus.

3.4.5 Import Users and Groups using the Security Update Feature

To import a user through the *Security Update* option, complete the following steps:

1. On the **Options** menu, click **Users, Groups and Roles**.
2. Enable the **Allow Database authentication** option on the **Users** tab.
3. Add the appropriate script to the project, as described in the *UpdateSystemSecurity* section of the *Oracle WebCenter Forms Recognition Scripting User's Guide*.
4. Configure RTS to execute **Update System Security** with this project.
5. Run RTS and wait for the update event.

3.4.6 Allow Database Authentication

Verifier uses the users and groups contained in the database to authenticate to the project. Therefore, complete the following steps for all projects:

1. On the **Options** menu, click **Users, Groups and Roles**.
2. In the **Project Authentication Properties** dialog box, on the **Users** tab, select **Allow Database Authentication**.
3. Verify that the users exist in the database.
4. On the **File** menu, click **Save Project**.

3.4.7 Export Users, Roles and Groups to the Forms Recognition Database

In Verifier, you have the ability to make the users, roles, and groups available in the WebCenter Forms Recognition database. This is necessary if a database platform is used by WebCenter Forms Recognition applications, like Web Verifier.

To make users, roles, and groups available in the database, use the **Export to Database** button in the project authentication dialog, which is available through the **Users, Groups and Roles...** option of the **Options** menus.

Note: Export only affects users and groups in the project being exported. Any other existing users or groups in the database are not affected.

3.5 Windows Based User Authentication

The Windows-based user authentication feature simplifies the log in procedure.

When you open a project that has the **Allow Windows Authentication** setting enabled in Designer or Verifier, the program tries to log in with the currently logged in Windows user's credentials without displaying the Oracle WebCenter Forms Recognition login dialog box.

If the currently logged in Windows user does not exist in the Oracle WebCenter Forms Recognition database, the Oracle WebCenter Forms Recognition login dialog displays. You cannot use your Windows user credentials in the Verifier and Designer login dialog box unless the password for the Windows user is explicitly set in the project settings and exported to the database.

3.5.1 Enable Windows Authentication

To enable Windows based authentication in **WebCenter Forms Recognition** for a desired project, open the corresponding project file in Designer, select the **Users, Groups and Roles...** from the **Options** menu, then select the **Allow Windows Authentication** check box on the **Users** tab of the *Project Authentication Settings* dialog.

You can enable Windows authentication and database authentication at the same time for different users.

3.5.2 Enable Windows Authentication for Web Verifier

To use Windows authentication for the Web Verifier login, complete the following steps:

1. Import the Windows accounts of the designated Web Verifier user to the project database through the **Windows Import** button or the *Security Update* method as described above.
2. Enable the Web server IIS Windows Authentication.

Note: Only domain users are able to login to the Web Verifier.

3.5.3 Import Windows Users

To import Windows users, you can use the “Windows Import” feature in Designer or the script based security update feature.

3.5.4 Import Windows Users in Designer

To import Windows users in Verifier, complete the following steps:

1. On the Options menu, click Users, Groups and Roles.
2. In the **Project Authentication Properties** dialog box, on the **Users** tab, select **Allow Windows Authentication** and then click **Windows Import**.
3. In the **Select Users or Groups** dialog box, click **Locations**.
4. In the **Locations** dialog box, select the domain that has the required accounts and then click **OK**.
5. In the **Select Users or Groups** dialog box, click **Advanced** and then click **Find Now**.
6. From the **Search results** list, select one or more users and then click **OK**.
7. In the **Select Users or Groups** dialog box, to import the users into the project, click **OK**.

Note: Verifier assigns all newly imported users to the predefined **Standard User** group. This group has the **SLV**, **VER**, and **FLT** roles assigned by default. You can modify this group according to the project requirements.

4 Modes

4.1 What Are Modes?

WebCenter Forms Recognition Designer supports several modes of operation:

- *Document Selection Mode*
- *Definition Mode*
- *Normal Train Mode*
- *Verifier Train Mode*
- *Runtime Mode*
- *Verifier Design Mode*
- *Verifier Test Mode*

Each mode supports a particular set of tasks used during the design of a WebCenter Forms Recognition application. When you switch from one mode to another, the toolbar and the window below it change to provide the tools and controls used in the new mode.

4.2 Document Selection Mode

4.2.1 Purpose

Use this mode for the following purposes.

- Import documents into new batches.
- Modify and open existing batches.
- View and modify classification and extraction learnsets.

4.2.2 Selection

To switch to *Document Selection Mode*, select one of the following options:

- From the **View** menu, select the **Document Selection Mode** option.
- On the toolbar, click the arrow beside the **Switch to Document Input Selection** button, then select the appropriate option (**Directory**, **Learnset**, or **Batch**) from the drop-down menu. Depending on the current input type, the appearance of this button varies.

If the paths for the *Batch Root* and *Image Root* are not correctly configured, the **Batch** menu option is disabled.

If you change the input type from the drop-down menu, *Document Selection Mode* is activated automatically.

4.2.3 Specify the LearnSet Location

To specify the location of the learnset, complete the following steps:

1. On the **Options** menu, select **Settings...**

2. On the *Settings* dialog, select the **Train Mode** tab.
3. For *Learnset Manager*, enter or browse to a **Base Directory**. To create a new directory, click the browse (...) button, then click **New Directory** and the path name.

4.2.4 Specify the Input Mode

To select an input type, complete the following steps:

1. On the **Options** menu, select **Settings**.
2. On the *Settings* dialog, select **Use Database as Documents and Statistics Source**.
3. On the **Input Mode** tab.
4. Specify the file format under **Import Options**.
 - You can select **Image & Electronic Documents** and enter the extension of the image file format, for example *.tif or similar.
 - You can also use workdocs without image files. In this case, select the **Workdoc Files (*.wdc)** option.
5. Optional. To create a new job, complete the following sub-steps:
 - a. Click the **Create Job** button.
 - b. Enter a name for the new job on the *Create Job* dialog.
6. Select the required job from the **Select Job** list.
7. Optional. Select **Save Execution Results to Workdoc** if you want to permanently save the result of the processing done in Verifier to the workdoc.

4.2.5 Specify the Location of Batches

This option is no longer in use.

To specify the location of batches that are stored on the file system, complete the following steps:

1. On the **Options** menu, select **Settings...**
2. On the *Settings* dialog, select the **Input Mode** tab.
3. Specify the location of the batch files in the **Batch-Root Directory** field.
4. For each batch file, there is a subdirectory containing the image files that will be processed. Specify the root of these subdirectories in the **Image-Root Directory** field. Typically, both root directories should be the same.

Note: You must define the batch location before you can select batches as input type.

4.2.6 Input Type Directory

Use *Directories* input type only if the required workdocs are already available. If workdocs are not available, WebCenter Forms Recognition creates new workdocs for the *TIFF* files.

Directory Pane	The top left pane displays the file system directories. To display the directory content in the bottom left pane, double-click a directory.
File list pane	The bottom left pane displays the files from the selected directory. To display a file in the right pane, click an entry.
Document Pane	The right pane displays a page of the currently selected document.

4.2.7 Create a Batch from Files in a Directory

To create batches from a directory, complete the following steps:

1. Select a directory from the directory list.
2. Right-click the directory and select **Import New Batch from Directory** from the popup menu. WebCenter Forms Recognition inserts the selected batch in the defined batch directory or database.

4.2.8 Input Type Batch

Runtime Server creates and processes batches. For classification and extraction design, use the *Batch* input type for *Document Selection Mode*, the user interface displays the following:

Batch Pane	The top left pane displays the batches from the batch root. To open the batch and display its structure in the bottom left pane, double click a batch.
Folder and workdoc pane	The bottom left pane displays the folders and workdocs of the selected batch. To display a workdoc in the right pane, select an entry.
Document pane	The right pane displays a page of the currently selected document.

You have several options for working with batches and documents in *Input Mode*. For batches, you can complete any of the following actions:

- Create a batch
- Rename a batch
- Change a batch state
- Change a batch priority
- Close a batch
- Delete a batch
- Add a folder

For folders, you can complete any of the following options:

- Rename a folder.

- Delete a folder and documents from the list.
- Add document.

For documents, you can complete any of the following options:

- Change state of a document to a new value that you assign.
- Delete the selected workdoc from the batch.

To select a batch option, right-click the desired batch and select the desired option from the popup menu.

To select a folder option, right-click the desired folder and select the desired option from the popup menu.

To select a document option, right-click the desired document and select the desired option from the popup menu.

You can also drag a folder to another batch or move a selected document to another folder within the same batch.

4.2.9 Input Type Learnset

To train classification and extraction, you need a Learnset. Select *Learnset* as input type to review your sample documents.

Document Classes Pane The top left pane displays the learnset classes. To display the documents of the class in the bottom left pane, select a class.

Classification/Extraction Pane The bottom left pane displays the documents of the selected class. To display a document in the right pane, select an entry.

Document pane The right pane displays the currently selected document.

4.2.9.1 Display the Classification LearnSet

The classification learnset is a set of sample documents used for classification. Right-click the document class and then **View Classification learnset**.

4.2.9.2 Display the Extraction LearnSet

To display the extraction learnset, Right-click the document class, and then click **View Extraction learnset**.

4.2.9.3 Remove a Document from the LearnSet

To remove a document from the learnset, complete the following steps:

1. In the bottom left pane, select a document.
2. Right-click the document and then click **Remove from LearnSet**.
3. Re-learn the documents.

4.2.9.4 Add a Document to an Engine

To add a document to the document class learnset of an engine, complete the following steps:

1. In the bottom left pane, select the required check boxes.
2. Re-learn the documents.

4.2.10 Automatic Document Conversion during Import

Converting files to a common image format, such as TIFF, on import may enhance and stabilize subsequent processing of these documents. WebCenter Forms Recognition provides the possibility to convert the imported electronic documents also. While it can be advantageous to process electronic documents directly, as electronic text may be extracted more quickly and more accurately, there are cases when conversion on import of such documents provides better results.

Advantages of automatic conversion

- The standardizing of document resolution stabilizes the document processing: generally set to 300 x 300 DPI for all learnsets and extracted images.
- In some cases, OCR post-processing of certain types of electronic documents may produce better extraction quality when it is image based.
- A single document format may be needed for archiving purposes, for example to archive original documents in an external database along with the extracted data entries in one unified format.

During import, WebCenter Forms Recognition converts each original document to the specified image file format attached to the associated workdoc. If the specified format supports multi-page files, all pages of the document are stored in a single file. Otherwise, each page of the document is saved as a separate image file and attached to a single workdoc.

On the **Processing** tab of the *Settings* dialog in *Runtime Server and Designer Import Option*, it is possible to configure a couple of options that allows automatic conversion of documents during import into the WebCenter Forms Recognition system. These options affect the following import processes.

- Import of documents in Verifier
- Import of documents using Runtime Server
- Handling of imported PDF documents during OCR processing

4.2.11 Automatic Document Conversion Settings

Convert Images into...	Select this option to specify the image format to which Verifier and Runtime Server convert the imported files. The default and recommended image format is TIF G4 compr, which supports multiple pages in one file.
<hr/>	
Automatically Convert Resolution to...	Use this option to specify the resolution for the converted files. The default and recommended value is 300 * 300.
<hr/>	
Convert Electronic Documents into Images	Select this option to convert CI-documents such as Excel, Word, Power Point, and non-image PDF files. The OCR process captures all data in the file including embedded images.
<hr/>	

Force OCR PDF as image

Select this option to retrieve image data from PDF documents with mixed content and at the same time retrieve electronic data when performing OCR on other types of electronic documents.

This option affects the OCR of PDF documents that have been imported without conversion to an image. By default, during OCR WebCenter Forms Recognition extracts electronic text from each page of a PDF document, ignoring any images that may be on the page. However, if a page contains no electronic text, the product performs OCR on the full page image, using project OCR engine settings and ignores any electronic text.

Note: This option is not available when you select "Convert electronic documents into images".

4.3 About the Definition Mode

In *Definition Mode*, following tasks can be performed:

- Design the project settings
- Establish classification and data extraction
- Generate a default export file
- Execute and export during classification and extraction
- Test single documents or an entire document set

4.4 Switch to Definition Mode

To switch to *Definition Mode*, perform one of the following options:

- From the **View** menu, select **Definition Mode**.
- On the toolbar, click the **Switch to Definition Mode** button.

4.5 Modify the Processing Delay

In *Definition Mode*, you can automatically process an entire document set in a single run. To do this, you will need enough time to view the results for one document before the next one displays.

To set the processing delay, complete the following steps:

1. From the **Options** menu, select **Settings...** or, in the toolbar, click the **Show Settings** button.
2. In the *Settings* dialog, select the **Definition Mode** tab.
3. Under **Definition Mode**, use the slider control to set the number of seconds delay, or, alternatively, enter the desired number of seconds into the associated text field.

4.6 Support of Non-Western Languages

Configure the Phrase Classification engine, Table Analysis engine and Format Analysis engine using phrases or strings in the desired non-western language on the corresponding non-computer or on any computer where this language is set as the system's default Unicode language.

This is also applicable for all other WebCenter Forms Recognition engines that do not require any visual input in non-native languages like the Brainware Extraction engine, the Brainware and ASSA Classification engines, the Brainware Layout Classification engine.

After support for non-western languages is activated, the project has to be saved and re-learned.

While the support option for non-western languages always looks for a one-to-one character transformation, the Chinese, Japanese and Korean language support may lead to an increase in text length of the transliterated object. The multi-character representation is phonetically more precise but increases the original text length. The attributes of the *SCBCroWorktext* objects like *TextLength* refer to the original character length. The attribute *TextProperty* and the WinWrap scripting engine's *Len* method return the length of the transliterated text. This limitation applies only during the transliteration phase, which usually corresponds to internal core classification and extraction and also involves some script events, like *PostEvaluate*, executed during the core extraction.

To use a language other than English, change the region and language settings of your system.

WebCenter Forms Recognition provides a generic approach to the support of "non-western" languages, where essential parts of the alphabet are not covered by letters of the English alphabet with the same key codes.

This includes the following languages:

- Greek
- Polish
- Chinese, Japanese, Korean and Taiwanese (CJKT)

These processes can be executed on a non-native computer, such as a United Kingdom (UK) or United States (US) workstation or server, to process documents that contain auxiliary information and information that is extracted in supported non-western languages.

The following WebCenter Forms Recognition engines allow phrases, format strings, and more to be entered with visual input using the characters of the currently supported non-western languages:

- Phrase Classification engine
- Table Analysis engine
- Format Analysis engine
- Associative Search engine (this engine includes the ability to import Unicode CSV files)

4.6.1 Byte Encoding for CJKT Languages

If you use Chinese, Japanese or Korean languages in your project, you need to enable the **Use Multi-Byte Encoding** option on the **Definition Mode** tab of the project's *Settings* dialog.

4.6.2 Advanced Options for Language Conversion

Use the `LanguageSupportWorkflowSettingsVisible` registry key to enable or disable advanced options for additional language conversion settings in Verifier. The settings provide features to convert words to extended ASCII character sets. To create the registry key, complete the following steps:

1. Launch the Windows registry editor.
2. Complete one of the following sub steps:
 - For a 32-bit machine, navigate to [HKEY_LOCAL_MACHINE\SOFTWARE\Oracle\Cedar].
 - For a 64-bit machine, navigate to [HKEY_LOCAL_MACHINE\SOFTWARE\Wow6432Node\Oracle\Cedar].
3. In the right pane, right-click and then click **New > DWORD (32-bit) Value**.
4. In the Name field, type **LanguageSupportWorkflowSettingsVisible** and then click **OK**.
5. Right-click the **LeaveDownloadedMessagesInInbox** key and then click **Modify**.
6. In the **Edit DWORD (32-bit) Value** dialog box, in the **Value data** field, type either of the following:
 - to disable, type **0** and then click **OK**.
 - to enable, type **1** and then click **OK**.

4.6.3 Copy Non-Western Language Phrase

If you are not able to type a non-western language phrase, complete the following steps to copy the phrase to the clipboard:

1. Switch to **Document Input Selection Mode** and load the required document.
2. Switch to **Verifier Train Mode**.
3. On the document, select the required phrase and copy it to a field.
4. From the field, copy the phrase to the clipboard.

4.6.4 Definition Mode User Interface

The user interface of *Definition Mode* is the core of WebCenter Forms Recognition Designer. The three distinct areas of the user interface are as shown below:

Class / Fields Panel	Displays the project classes. The Fields tab displays the fields that Verifier extracts from the documents.
Document Panel	Displays the most recently selected document.
Properties Panel	Displays the properties for the select project, class or field. To display the properties panel, on the toolbar, click the Show/Hide class and field properties button.

4.6.5 Project-Level Properties

To display the project properties, complete the following steps.

1. In the left panel, on the **Classes** tab, select the project.
2. If the properties do not display in the right panel, select the project and then click the **Show/Hide class and field properties** button.

4.6.5.1 Project tab

The Project tab contains general classification settings such as classification interpretation, threshold, and distance for standard and parent classification, OCR settings, word segmentation characters, and storage options.

4.6.5.2 Classification tab

The Classification tab contains the classification engines and their properties as well as default classification class.

4.6.5.3 Advanced tab

The Advanced tab contains options to force validation, remove preceding and trailing spaces, and show custom class names and to mark only one field when the validation rule fails.

4.6.5.4 Supervised Learning tab

Supervised Learning is configurable only at the project level. Refer to [Section 7.15: ASSA Classify Engine](#) and [Section 12: Setup Supervised Learning](#) for information about configuring Supervised Learning.

4.6.6 Class-Level Properties

Three types of project-level properties are available for definition when you select a base class or derived class from the **Class** tab on the left side of the screen. These are described below:

Classification tab	The Classification tab contains the available classification engines and their properties.
Document Class tab	The Document Class tab contains the properties of the selected document class.
Validation tab	The Validation tab contains the validation settings.

4.6.7 Display Field Properties

To display the field properties, complete the following steps.

1. In the left panel, on the **Fields** tab, select a field.
2. If the properties do not display in the right panel, right-click the field and then click **Show Properties**.

4.6.7.1 Field-Level Properties

Four types of field-level properties are available, these are described below:

Analysis tab	This tab contains the available templates, the available analysis and their settings.
Evaluation tab	This tab contains the available evaluation engines and their settings.
Field tab	This tab contains the OCR settings for the selected field.
Validation tab	The tab contains the available validation engines and their settings.

4.7 Normal Train Mode

4.7.1 About Normal Train Mode

This mode enables you to train classification and extraction using the properties defined in *Definition Mode* and the current document set selected in *Document Selection Mode*. Although learning is also available in *Definition Mode*, only *Train Mode* allows the creation of learnsets.

4.7.2 Switch to Normal Train Mode

To switch to *Normal Train Mode*, complete the following step:

1. From the **View** menu, select the **Train Mode** option. Alternatively, click the arrow next to the **Switch to Train Mode** toolbar button, then select **Normal Train Mode** from the drop-down menu.

4.7.3 Specify Train Mode Settings

Train Mode requires particular settings that enable you to create a learnset for classification and data extraction.

To set the options for *Train Mode*, complete the following steps:

1. From the **Options** menu, select **Settings...** Alternatively, click the **Show Settings** toolbar button.
2. In the *Settings* dialog, select the **Train Mode** tab.
3. To create a learnset or add documents to it, check the **Add Trained Documents to the Learnset** check box. Clear the check box only if you want to use *Train Mode* for demonstration purposes. By default, the check box is checked when you create a new project.
4. Select the **Execute Classification for New Documents Only** check box if you want Designer to classify each new document that is to be added to the learnset for data extraction. You can use this option if your sample documents do not belong to the same classes. Clear the check box if you work with sorted input.
5. If **Review in Normal Train Mode Previously Extracted Values** is activated, during learning in *Normal Train Mode* the system will use previously extracted values from the document and pre-fill form fields with them.
6. Under **Learnset Manager**, specify a directory where the documents of the learnset will be stored.

Note: Every time you add a sample document to the learnset of a class, Verifier copies the document to this directory.

7. Optional. To create a backup before learning new documents, select **Always automatically create a learnset backup before applying learning** check box.

4.7.4 Train Mode User Interface

The user interface of *Train Mode* enables you to train the classification and extraction learnsets.

Class Correction	<p>To add a document automatically to the classification learnset, select the Classes tab.</p> <p>In Classes tab you can complete the following options:</p> <ul style="list-style-type: none">▪ Manually assign a document to a class.▪ Modify the proposed Class assignments. <p>To add a document automatically to the extraction learnset, select a field on the Fields tab.</p> <p>In Fields tab you can complete the following options:</p> <ul style="list-style-type: none">▪ Manually select the correct field for data abstraction from several candidates.▪ Modify the proposed candidates for data extraction.
Document View	Displays the currently selected document.

4.8 Verifier Train Mode

4.8.1 About Verifier Train Mode

Use the Verifier Train Mode to check the extraction result and for Automatic Supervised Learning. To do this, specify the settings for the project level in Definition Mode on the Supervised Learning editor and on the **DocClass** level on the **Document Class** tab.

4.8.2 Automatic Supervised Learning

Automatic Supervised Learning is a Verifier extension that uses Associative Learning to extract supplier information. Automatic Supervised Learning improves the global knowledge base, the quality of documents, and decreases the time to design applications.

When a supplier is extracted correctly and this document is added to the learnset, Verifier creates a new class automatically.

Automatic Supervised Learning then:

- Extracts the supplier information.
- Creates a new class using the supplier name if the document added to the learnset.
- Learns the new class.

- Migrates the local project to the global learnset.

4.8.3 Enable Verifier Train Mode

To enable the *Verifier Train Mode*, complete the following step:

1. From the **View** menu, select the **Verifier Train Mode** option. Alternatively, click the arrow next to the **Switch to Train Mode** toolbar button, then select **Verifier Train Mode** from the drop-down menu. Select **Enable Automatic Supervised Learning Workflow for this Project**.
2. Create a verification form.

4.8.4 Verifier Train Mode Settings

To define the settings, complete the following steps:

1. From the **Options** menu, select **Settings...** Alternatively, click the **Show settings** toolbar button.
To create a new classes using the supplier's name and to create a learnset, select the **Apply Local Classification and Extraction Automatically** check box.
2. Select the **Prompt if Script Forces or Rejects Insertion to the Learnset** check box.
3. Under **Put Document to Local Learnset**, select one of the following options:

Only if Adding Activated by a User	Verify that a document was added to the learnset if the document count is updated for a class created as the result of supplier extraction.
Automatically if More Than...	Learn documents only when they are initially unclassified.
Always Prompt Before Adding	If checked, the Verifier user will be prompted before a document is automatically added to the local learnset.

4. Under **Learn New Documents**, select one of the following options:

Only by User Request	Initiate learning only if the user requests it.
Before Batch Closing	Initiate learning for every batch in the project each time any batch is closed.
Immediately	Initiate learning any time a document is added to the learnset.

5. Select Save all path information relative to the project file and then click **OK**.

4.8.5 Switch to Verifier Train Mode

To switch to Verifier Train Mode, complete the following step:

- On the toolbar, from the Switch to **Train Mode** list, select **Verifier Train Mode**.

4.8.6 Train the Learnset

To train the learn set, complete the following steps:










1. Open a document.
2. On the toolbar, from the **Switch to Train Mode** list, select **Verifier Train Mode**.
3. Click the **Classify / analyze current document** button.
4. Ensure that the **Add document to Learnset** button is active.






4.8.7 Verifier Train Mode User Interface

The four areas of the *Verifier Train Mode* user interface are described below:

Form Area	<p>The form area contains a set of controls for correction and manual indexing.</p> <p>Each control is assigned to a data extraction field. The viewers show the document areas that were extracted to fill the fields.</p>
Current Input Area	<p>This area shows a magnification of the current form field and the associated viewer. It can be used for particularly convenient viewing and editing.</p> <p>Note that the position of the form areas may vary according to the selected layout, or only a subset of form areas may be present.</p>
User Info Area	<p>This area displays the class name of the current document, the name of the current field and, if applicable, a message indicating why the current field is invalid.</p>
Document Area	<p>This area displays the current document.</p>

4.8.8 Verifier Train Mode Toolbar

Button	Description
	Moves to the previous page
	Moves to the next page
	Highlight all fields
	Highlight selected fields
	Highlight all candidates
	Maximize image height
	Maximize image width
	Zooms in.
	Zooms out.

	Keep focus on field
	Keep zoom on switching to next document
	Move to next document on validate
	Force document to be validated with selected form
	Show/hide script

4.9 Runtime Mode

4.9.1 About the Runtime Mode

Use this mode to perform the following tasks:

- Simulate classification and extraction using the entire document set.
- Test the export scripts.
- Perform OCR on a document set that will be used in *Definition Mode* or *Train Mode*.

Note: Except for the OCR output and the corresponding images, the results of the *Runtime Mode* are not saved persistently.

4.9.2 Switch to Runtime Mode

To switch to *Runtime Mode*, complete the following step:

1. From the **View** menu, select the **Runtime Mode** option. Alternatively, click the **Switch to Runtime Mode** toolbar button.

4.9.3 Settings

For *Runtime Mode*, you need to define import and export options. In addition, you can either test classification only or classification plus extraction.

To set the options for *Runtime Mode*, complete the following steps:

1. From the **Options** menu, select **Settings**. Alternatively, click the **Show Settings** toolbar button. The *Settings* dialog is displayed.
2. On the *Settings* dialog, click the **Runtime Mode** tab.
3. Under **Workflow Settings**, select the desired options from the following:

Execute Classification	Execute the classification of input automatically.
-------------------------------	--

Redo OCR on Documents	Rerun the OCR process with selected settings.
------------------------------	---

Execute Extraction	Execute the extraction of input automatically.
---------------------------	--

Execute Export	Execute the export of documents automatically.
-----------------------	--

- Under **Export Options for Definition and Runtime Mode**, select the desired options from the following:

Trigger Script Based Export	Start a predefined script that contains export instructions.
Copy Documents to Export Directory	Copy processed documents from the import to the export directory instead of moving them.
Generate Default Export File for Each Batch	Generate an ASCII file with the processing results. The export file will be recreated with each run. It will be written to the export directory. The file name is generated according to the naming convention <i><name of the import directory>.exp</i> .
Generate PDF File on Export	Generate a PDF file with the processing results. The export file will be recreated with each run and written to the export directory. The file name is generated according to the naming convention <i><name of the import directory>.pdf</i> .
Export Directory	Specify a directory location where images and OCR output are stored after processing.

4.9.4 About Parent Classes

Parent classes sometimes contain multiple derived classes that might also be parent classes themselves. If a document class is derived from one or more parent classes, each event for that class will also be executed in all parent document classes, starting at the most senior parent class.

Both *bottom up* and *top down* apply to available events.

For most users, the default setting of **Top Down** is the most appropriated. Change the settings for inheritance chronology event sequences only if there are scripts created for those events. However, unless there is a lot of code, it is probably easier to change the code, rather than change the event chronology.

4.9.5 Compatibility Settings

On the *Settings* dialog, there is an additional tab associated with the *Runtime Mode*; the **Compatibility** tab.

The following settings are available:

- Under *Inheritance Chronology - Event Sequence*, check or clear the **Top-Down (From Parent to Derived Class)** check box as appropriate.
 - When this option is checked, the inheritance succession goes from the parent class to derived classes.
- When it is cleared, the inheritance succession is reversed and goes from the derived class to the parent class. Under **Advanced Validation Option**, select the desired options from the following:
 - Do Not Allow Script Overwriting of Standard Validation Output.

- Do Not Fire Validation Events for a Class if 'Use Derived Validation' Flag Is Set.
- In Verifier Validate Restricted Characters in All Cases and Not Only When Typing.
- Fire 'OnAction' Script Event Recursively.
- Allow Firing of FocusChanged Event When Loading the Verification Form.
- Allow Firing of VerifierFormLoad even when in Verifier Test/Train Modes.

4.9.5.1 User Interface

The four areas of the *Runtime Mode* user interface are described below:

Statistics View	The Statistics View displays the classification and extraction results in a hierarchical representation of the classes defined in the current project. The view updates continuously at runtime.
Class Selection	The class selection area contains a hierarchical representation of classes that are defined in the current project.
Classification Results	Displays a list of all documents processed in the current batch and the classification result for each document.
Document Area	This area displays the current document.
Field Extraction Results	This area displays the fields defined for the class that the document belongs to, and the extracted values for each field for the current document.

4.10 Verifier Design Mode

4.10.1 About Verifier Design Mode

In *Verifier Design Mode*, you create verification forms that you can use in Verifier to manually correct extracted data and for manual indexing. Switch to Verifier Design Mode

To switch to *Verifier Design Mode*, complete the following step:

1. From the **View** menu, select the **Verifier Design Mode** option. Alternatively, click the arrow next to the **Switch to Verifier Mode** toolbar button, then select **Verifier Design Mode** from the drop-down menu.

4.11 Verifier Test Mode

4.11.1 About Verifier Test Mode

The *Verifier Test Mode* simulates the behavior of Verifier, use this mode to test the verification with forms that were created in *Verifier Design Mode*.

4.11.2 Switch to Verifier Test Mode

To switch to *Verifier Test Mode*, complete the following step:

1. From the **View** menu, select the **Verifier Test Mode** option. Alternatively, click the arrow next to the **Switch to Verifier Mode** toolbar button, then select **Verifier Test Mode** from the drop-down menu.

5 Working with Documents

5.1.1 What Are Workdocs?

Starting with the image or file that is provided as input, WebCenter Forms Recognition processes the document and writes the processing result to a structure called a *workdoc*.

The workdoc is created during the initial OCR of an image or the initial filtering of a file. The OCR also includes a layout analysis that identifies words and blocks within the document. All subsequent processing steps use the workdoc and not the original file. While the document is being processed, each step adds its results to the workdoc. This includes the following results:






- The document's text.
- The location of words.
- The geometrical relationship between words.
- The classification results.
- The fields for data extraction.
- The candidates for field contents.
- The selected correct candidate for each field.
- The fields' validation status.

The process of analyzing a document is complete when the workdoc is assigned to a document class and all the fields for extraction are filled and valid.










5.1.2 View a Document

After you select the document input, the corresponding set of documents displays in the Designer viewer. The viewer displays the image or file, and structures in the workdoc.

The application's status bar displays the total number of documents in the current set and the serial number of the current document. A navigation bar in the toolbar enables you to navigate through the document set:

Button	Description
	Move to the first document.
	Move to the previous document.
	Move to the next document.
	Move to the last document.
	Displays the serial number of the current document. To move to a particular document, type its number in this field and press the [Enter] key.

The document viewer's toolbar provides access to zooming functions and enables you to navigate through multi-page documents.

Button	Description
	Adjusts the document display to the height of the viewer window.
	Adjusts the document display to the width of the viewer window.
	Adjusts the document display to the height or the width of the viewer window so that the entire page displays.
	Zooms in.
	Zooms out.
	Moves to the first page of a multi-page document.
	Moves to the previous page of a multi-page document.
	Moves to the next page of a multi-page document.
	Moves to the last page of a multi-page document.

5.1.3 Switch between Workdoc and Page Definition View

In *Definition Mode*, you can display either the scanned image or the workdoc. By default, the workdoc is displayed. To create reading zones on a document, you need to work with the image. Designer automatically displays the image when you start creating zones.

Use the **View** menu between image display and workdoc display to view an image or to display the workdoc.

- Select **Show Page** to display the image.
- Select **Show Workdoc** to display the workdoc.

5.1.4 Process a Document

The process selection group enables you to process single or multiple documents. Depending on the state of the program, processing may include the following features:

- OCR (if there is no workdoc for the current document).
- Classification (if there are classes).
- Data extraction (if there are fields; in *Runtime Mode* or *Train Mode*, the **Fields** tab must also be active).
- Document Export (only available in definition and *Runtime Mode*).
- Verification (only available in *Verifier Test Mode*).








5.1.5 About Learning

The learning process performs classification and data extraction. For classification, learning requires that you provide examples that indicate which documents belong to a specified class. For extraction, learning requires that you select the correct data from a set of extraction candidates.

WebCenter Forms Recognition Designer takes the examples you provide and creates learnsets from them. It then learns to produce the correct output from input that is similar to the learnset.

Use the *Learning* toolbar to create learnsets and trigger the learning.

5.1.6 Learning Toolbar

Button	Description
	Processes the current document. The button's drop-down menu enables or disables the debug mode.
	Processes the next document.
	Processes documents in the current set starting with the current one.
	Pauses the processing of documents.
	Stops the processing of documents. Clears any results from previous runs.
	Adds the current document to the Learnset of the selected class.
	Starts the learning for the current project.

5.1.7 Learning Toolbar

Learning is involved in classification and in data extraction. For classification, learning requires that you provide examples that indicate which documents belong to a specified class. For extraction, learning requires that you select the correct data from a set of extraction candidates.

WebCenter Forms Recognition Designer takes the examples you provide and creates learnsets from them. It then learns to produce the correct output from input that is similar to the learnset.

Use the *Learning* toolbar to create learnsets and trigger the learning.

5.1.8 Compress a Learnset

WebCenter Forms Recognition Designer saves workdoc files in a compressed format. To update a learnset with changes to the *FieldID*, complete the following steps:

1. Switch to *Definition Mode*, *Train Mode*, or *Verifier Train Mode*.
2. From the Options menu, select **Compress Learnset**.
3. Click **Yes** to confirm the action.

5.1.9 About Highlighting






Highlighting visualizes structures that have been created during document processing. Therefore, this feature is only available when a workdoc exists. In addition, field and candidate highlighting is only available when fields are defined.

If table analysis is used, only the best table candidate can be highlighted. Table highlighting must not be triggered using toolbar buttons but is active by default. The behavior varies with the different modes:

In *Definition Mode*, table elements are highlighted during the definition process, and the entire table is highlighted when a document has been processed. Valid elements are highlighted in green. Invalid elements are red.

In *Verifier Test Mode*, the extracted table is available and can be used to control which elements of the document table are highlighted. Valid elements are highlighted in green. Invalid elements are red.

5.1.10 Highlighting Toolbar

Button	Description	View Menu
	Displays the document without highlighting.	Highlight Nothing
	Highlights all words in the document in turquoise. Words are a result of the OCR process. Each character is stored together with its position. To view the OCR result, point to a word with the mouse. The recognized word displays as a tooltip.	Highlight Words
	Highlights all blocks in the document in yellow. Blocks are groups of words identified in <i>Definition Mode</i> by their geometrical relationship.	Highlight Blocks
	Highlights all candidates for the currently selected field in maroon. A candidate is a possible value for a field identified in the processing step. To view the weight of the candidate, point to it with the mouse. The weight displays as a tooltip.	Highlight Candidates
	Highlights all fields in the document. If the field is valid it is highlighted in green, otherwise it is highlighted in red. To view the name of the field, point to it with the mouse. The name displays as a tooltip.	Highlight Fields
NA	Only relevant in conjunction with table extraction or scripts using the workdoc API. Highlights all words where Visible = TRUE.	Highlight Checked Words

5.1.11 Print documents

You can print documents including pages with highlights. Complete the following steps:

1. On the **File** menu, click **Print**.
2. In the **Print** dialog box, click **Print**.

6 Working with Projects

6.1.1 What Are Project Files?

In a WebCenter Forms Recognition project, the automatic analysis and processing of incoming documents relies on a group of settings and on trained knowledge. Verifier saves these settings persistently in a project file. The project file stores the following information:

- General settings such as the location of import and export directories.
- Document classes and their hierarchical relationship.
- References to learnsets for full-text classification and their status.
- Phrases and rules for rule-based classification.
- Settings for image size classification.
- Field names and field positions for data extraction.
- Analysis settings and standard validation rules for data extraction.
- Field recognition settings for data extraction.
- Table recognition settings for data extraction.
- Neural networks for full-text classification.
- Neural networks for data extraction and their status.
- Form definitions for data verification in WebCenter Forms Recognition Verifier.
- Scripts for document classification, data extraction, data validation, data verification, and document export.

6.1.2 Open a Project

If WebCenter Forms Recognition Designer is active, there are four ways to open a project:

- From the **File** menu, select the **Load Project...** command, then browse to and select a project (.sdp) file from the file system and open it.
- From the **File** menu, select one of the recent projects. These are listed at the bottom of the menu.
- On the toolbar, click the **Load Project** button, and then select an SDP file from the file system and open it.
- Run the Designer executable from a command line and specify the full path and filename for the project to open as a command line argument.

You can load or save a file with or without learn data. When loading a project, click the **File Type** drop-down box and select either **WebCenter Forms Recognition Project (*.sdp)** or **WebCenter Forms Recognition Project, Skip Learn Data (*.sdp)**.

6.1.3 Create a Project

In WebCenter Forms Recognition Designer, there is always precisely one active project. There are two ways to create a project, depending on the application's state.

If Designer is active:

- From the **File** menu, select the **New Project** option, then follow the instructions on screen.

If Designer is not active or to start a new instance:

- Start Designer through the Windows Start menu. This creates a new project automatically.

6.1.4 Save a Project

When you have modified your current project and attempt to close it, Designer asks you to save changes that could be lost. If you want to save a project without closing it, use one of the following methods.

- On the **File** menu, select **Save Project**. This saves an existing project file. If your project is new, you need to enter a file name. The project file is then saved with the extension **.sdp**.
- Click the **Save Project** button on the toolbar.

You can also save your project under a new name. Select the **File** menu and then the **Save Project As...** option. You need to specify a new file name.

Save and Compress is a third option available for saving your project. This functionality prevents the growth of a project file that is saved after a document class is removed.

Note: Neither the project file name, nor any folder names in its directory path, can contain any multi-byte characters.

6.1.5 About the GUI character set

Designer does not support Unicode, especially for elements in the user interface.

When configuring the project, for example when creating classes, only use characters of the English character set.

6.1.6 Save a Project without Learn Data

You can save a project without learn data. To select, in the *Save As* dialog, select the file type drop-down, then select either **<project name> (SDP)** or **<project name> Without Learn Data (SDP)**.

6.1.7 Save Compress Project

This feature is needed because saving a project file after removing a document class and its associated learnset causes the project file to grow or remain the same rather than to decrease.

Choosing **Save and Compress** allows you to complete the following options:

- Save a temporary project file.
- Replace the original file with this temporary file.
- Remove the temporary project file.

Because **Save and Compress** is slower than **Save** or **Save Project As...**, you should only use this option when you release a project file for production, not when you are actively working on the project.

However, using **Save and Compress** does not cause the size of the project file to decrease from the size it was before the document class was deleted. Rather, it keeps the size the same as it was before the deletion. To actually decrease the size of the file, you need to use the **Save Project As...** option.

6.1.8 Project files Version Control

WebCenter Forms Recognition Designer features an integrated version control system. This feature is useful if you are designing complex applications. With version control, you can save your work or back it up in intermediate stages.

For example, you could design your application in stages. In the first stage, define the document classification. Once this part is working, create a version of your project. If anything goes wrong during the design of the extraction part, you can always go back to the version you created earlier.

By default, version control is off. You can activate it when you are about to save an intermediate stage.

To use version control, complete the following steps:

1. From the **File** menu, select the **Versions...** option.
2. In the *Versions* dialog box, complete the following actions:
 - To specify the maximum number of versions for the current project, enter an integer in the **History Depth** text field. Once the maximum number of versions is reached, you will not be able to create additional versions. In this case, increase the maximum number of versions or delete versions that are no longer required.
 - To create a new version, click **Save as New** and then enter a description for the new version.
 - To access a previously saved version of your project, select it from the list of versions and click **Open**.
 - To delete a previously saved version of your project, select it from the list of versions and click **Delete**.

6.1.9 Client Keys

6.1.9.1 About Client Keys

Client keys are available in the *Verifier Runtime Mode*, in *Runtime Server*, and in *Verifier* when you select the respective client. You can use the keys as project variables in scripts. It enables you to manage different sets of values for multiple clients or customers.

The client keys are organized in one hierarchical parent – child structure. This tree is unique for all clients, while the key values are specific for the selected client.

6.1.9.2 Display the Client Keys Tree

To display the client keys tree, complete the following steps:

1. On the **Options** menu, select **Settings...** Alternatively, on the toolbar, select the **Show Settings** button.
2. In the *Settings* dialog, select the **Client Keys** tab.
3. To display the client-specific values, from the **Active Client** list, select the required client.

6.1.9.3 Manage Clients

By default, each WebCenter Forms Recognition project contains a standard client named **Default**. If your project is to be used for more than one client, you can create additional clients.

To create a client, complete the following steps:

1. In the **Client Keys** tab, click **Add**.
2. On the *Add Client* dialog, enter a name for the client and click **OK**.
3. To select a client, in the **Client Keys** tab, select an entry from the **Active Client** list box.
4. To delete a client, in the **Client Keys** tab, click **Remove**. The active client is deleted when you confirm the resulting notification.
5. To delete all clients, including their associated keys, in the **Client Keys** tab, click the **Clear Settings** button. The entire settings tree is deleted when you confirm the resulting notification.

6.1.9.4 Manage Keys

You can create a hierarchical structure of keys, consisting of root nodes and child nodes. The same keys are used for all clients.

To add a root node to the hierarchy of keys, complete the following steps:

1. In the **Client Keys** tab, right-click within the area in the center of the dialog box and select **Add Root Key** from the popup menu. The *Add New Key* dialog is displayed.
2. Enter a unique name for the key and click **OK**. The new key is inserted at the top level of the hierarchy.

To add a child node to the hierarchy of keys, complete the following steps:

1. On the **Client Keys** tab, right-click an existing key and select **Add Child Key** from the popup menu.
2. In the *Add New Key* dialog box, enter a unique name for the key and confirm with **OK**. The new key is inserted below the parent key.

To remove a key, complete the following steps:

1. On the **Client Keys** tab, right-click an existing key and select **Remove Key** from the popup menu.
2. Confirm that you want to delete the key.

To specify a value to key, complete the following steps:

1. Select the **Client Keys** tab.
2. Select a client from the **Active Client** drop-down.
3. Click the key you want to modify.
4. In the **Value** field, type the value that needs to be associated with the key name specified in the **Key** column.
5. Press the [Enter] key to accept the new value.

6.1.9.5 Manage Values

An individual value can be specified for each defined key from the tree. This value is going to be used only for the currently selected *Active Client*.

1. Select the **Client Keys** tab.
2. Select a client from the **Active Client** drop-down.
3. Click the key you want to modify.
4. In the **Value** field, type the value that needs to be associated with the key name specified in the **Key** column.
5. Press the [Enter] key to accept the new value.

6.1.10 Absolute and Relative Directories

The project file is portable so that it can be easily reused. WebCenter Forms Recognition Designer can work either with absolute or with relative references to the following directories:

- Export directory
- Learnset directory

6.1.10.1 Enable Relative References

By default, the project uses absolute references. For full portability, you should use relative references.

To enable relative references:

1. From the **Options** menu, select **Settings...** Alternatively, click the **Show settings** toolbar button.
2. In the *Settings* dialog, on the **Verifier Mode** tab, check the **Save All Path Information Relative to the Project File** check box.

6.1.10.2 Copy Project

To transfer a project, copy the project (.sdp) file and, if applicable, the documents from the learnset to the target directory.

With absolute references, adjust the path settings in the project options:

1. On the **Options** menu, click **Settings**.
2. In the **Settings** dialog box, on the **Train Mode** tab, in the **Base Directory** field, type or browse to the required directory.
3. On the **Runtime Mode** tab, in the **Export Directory** field, type or browse to the required directory and then click **OK**.

6.1.11 Project Structure

This section illustrates a typical project's folder structure and the types of files and file extensions in use. Each project has its own folder structure. The root folder for the project carries the same name as the project name (the .sdp file you created for the project). Folder structures for projects that did not use associative learning differ from those that do use associative learning.

6.1.12 Non-Associative Projects

In non-associative projects, the project (*.sdp) file, templates and comma-delimited reference files (CSV), such as vendor pools, are stored at the project root.

Below the project root, all projects have batch folders, export folders and learn folders. The primary batch folder holds batch state files, batch control files, lock files, .sdb files and .sdl files. Batch folders contain secondary folders for each batch in the project. These folders hold the workdocs and image files for each document in the batch. The learn folder contains all the learnsets for the project. Each of these learnset folders contain the images and workdocs in the learnset. The export folder contains exported files.

6.1.13 Project Planning

In WebCenter Forms Recognition, when you plan to create a new project, follow this sequence. Each step requires a careful analysis of the current document input as well as the associated business processes.

1. Analyze the format of the document you want to import.
2. Optional. Set up document separation.
3. Analyze the OCR requirements.
4. Set up classification.
5. Set up data extraction.
6. Optional. Set up verification.
7. Set up evaluation.
8. Optional. Set up Supervised Learning.
9. Set up document export.

6.1.13.1 Document Import Formats

You can process scanned paper documents, faxes, files, or documents from mixed sources.

Paper documents are typically forwarded as TIFF images to WebCenter Forms Recognition as Black-and-white images with a resolution of 300 dpi work best. Although OCR optimization can be done in WebCenter Forms Recognition, there is no way to regain information that was lost during scanning.

To process and extract text from CI documents, such as Word, Excel, and ASCII files, WebCenter Forms Recognition's built-in filters are available. Each file format requires individual testing. You may need to convert the documents to a supported format before you import them into WebCenter Forms Recognition.

To process emails, you may need to save the messages to the file system first.

6.1.13.2 Document Classes

Use the following document properties to define the initial classification hierarchy:

- Similar textual content
- Reoccurring phrases
- Reoccurring forms or letterhead paper
- Characteristic document dimensions

You cannot define a classification hierarchy based on document size only.

6.1.13.3 Classification Engines

WebCenter Forms Recognition offers several classification engines that you can use stand-alone or in combination with each other.

The classification engines are specified at the class level. You can apply one or more classification engines to each class. An evaluation algorithm combines the results and decides whether a given document belongs to a particular class.

At this stage, you need only an idea of classification methods. To improve the initial classification results, you can easily add classification methods later.

You can use one or more of the following classification engines:

- Brainware Classify Engine
- Layout Classify Engine
- ASSA Classify Engine
- Brainware Layout Classification Engine
- Forms Classify Engine
- Phrase Classify Engine
- Language Classify Engine
- Image Size Classification Engine

6.1.13.4 Extraction engines

Extraction of information from the document into a field occurs in two distinct steps.

- **Analysis:** In this step, the document content is analyzed and a set of possible values for a field is generated. These values are called candidates.
- **Evaluation:** In this step, the correct candidate is selected from the set of candidates.

WebCenter Forms Recognition supports several analysis methods. The best choice for the evaluation method depends on the selected analysis method. Analysis and evaluation methods are defined at the field level.

Although it is possible to combine several classification methods for each class, you can use only one analysis method and one evaluation method per field.

Evaluation is normally only necessary for the Format Analysis engine. You can also use it for the Zone Analysis Engine if various zones are defined. Other engines do not need an evaluation engine.

The following extraction engines are available:

- Associative Search Engine
- Brainware Table Extraction
- Check Analysis Engine
- Format Analysis Engine
- Table Analysis Engine
- Zone Analysis Engine
- Line Extraction Engine

6.1.13.5 Planning the Verification

WebCenter Forms Recognition features a dedicated program for quality assurance: Perceptive Intelligent Capture Verifier. Verifier provides the means to check and correct uncertain or invalid results of automatic document classification and data extraction. In addition, it permits manual classification and indexing of documents.

For each class, decide which steps are to be carried out automatically, which steps are to be carried out manually, and which verification steps will be involved (Classification and extraction verification separately, or both procedures in one step). If manual indexing is involved, decide whether this should be done with or without database support (Smart indexing).

Verification can also be carried out in stages, and exception handling mechanisms can be implemented to account for special error situations.

Plan the verification forms that will be required and the validation rules that will be imposed on user input.

6.1.13.6 Planning Supervised Learning

The Supervised Learning Workflow (SLW) is used to automatically create and learn new document classes. This automatic process creates the new derived documents classes for a generic - base - document class (Level 0). This document class is not inherited from any other document class. The derived document classes (Level 1) cannot have further derived document classes.

6.1.13.7 Planning the Document Export

By default, WebCenter Forms Recognition uses a specified export directory to save documents and their associated workdocs after processing.

For more information, refer to [Section 4.9.3: Runtime Mode settings](#) and [Section 16.1: About setting up the document export](#).

6.1.13.8 Document Separation

Use this feature to merge single-page documents into multi-page documents. The feature is optimized for invoices.

For more information, refer to *Configure an Import Instance for Document Separation* in the *WebCenter Forms Recognition Runtime Server Guide*.

6.1.13.8.1 Prepare a LearnSet for Document Separation

The document separation feature needs its own learnset. To prepare a learnset for document separation, complete the following steps:

1. Create a new project.
2. Switch to **Definition Mode**.
3. In the left pane, on the **Classes** tab, right-click the project and then click Insert **Base DocClass**.
4. In the **Create New Base DocClass** dialog box, type a name for the class and then click **OK**.
5. In the left pane, on the **Fields** tab, insert the following fields in the given order. For more information, refer to [Section 6.1.13.8.2: Fields in the Document Separation Learnset](#):
 - a. InvoiceNumber
 - b. Date

- c. Page
 - d. PageNumber
 - e. SenderID
6. Prepare a directory that contains about 50 fully extracted and validated, single-sided documents with workdocs.
 7. Launch **Verifier**.
 8. Open the first document.
 9. Assign the fields using selection tool feature.

Note: If some fields are not present in the document, simply leave them empty.

10. Repeat the previous step for all documents.

6.1.13.8.2 Fields in Document Separation Learnset

InvoiceNumber	A principle identifier of a multi-page document. If the Invoice number changes between two pages, a new document will start.
Date	Another criterion that can strengthen the determination of a multi-page document.
Page	This field has to contain the page keyword to strengthen the page number extraction. For example, the <i>Page</i> keyword could be Page or Pg.Nr. or something similar; whatever is present on the document.
PageNumber	If the document has pagination, a new document can be recognized when the pagination starts over.
SenderID	The identifier of an invoice type. For example, the tax number of the vendor can be used to identify a unique invoice of multiple pages from the same vendor. If the SenderID changes between two pages, this signals that a new document has started.

6.1.13.9 Multipage Detection

Multipage detection uses phrases to find documents that belong together and makes multipage TIFF image documents from them. The documents are checked during the classification and extraction steps of batch processing. Multipage detection works only within each folder of the batch. You must scan the relevant documents in the correct order to the same folder.

Note: If pages of a document are spread to different folders, multipage detection will not work even when the folders belong to the same batch.

For more information, refer to *Configure an instance for multipage detection* in the *WebCenter Forms Recognition Runtime Server Guide*.

6.1.13.9.1 Activate Multipage Detection

To activate multipage detection, complete the following steps:

1. Ensure Designer is in *Definition Mode*.
2. On the left side of the screen, the **Classes** tab is activated.
3. On the **Input** tab for the project settings the option to **Save Execution Results to Workdoc** is enabled. Otherwise, the changes cannot be stored.
4. On the **Classes** tab, right-click the project node to display a popup menu.
5. Select **Show Multipage Detection** from the popup menu.
6. On the **Multipage Detection** properties tab, select the **Activate Multipage Detection** check box.
7. Select a default for the start of detection, either **First Page** or **Following Page**, from the drop-down list.
8. To keep the page sizes similar, select **Similar Page Size Required** and select the **Tolerance** for the page size.
9. Select the **Phrase Multipage Detection** tab.
10. Select the **Activate** check box.
11. Set the **Threshold** and **Distance** values, if desired, for **Multipage Detection**. The default is **60** percent for **Threshold**, and **20** percent for **Distance**.
12. Select **First Page**, **Subsequent Page**, and **Last Page** tabs to define phrases for the different pages as appropriate.
13. To enter a phrase, edit the **New Phrase** text field.

Note: The maximum character length for a phrase is 32.

14. Click the **Ins** button.
 - Optional. From the **Position** list, select a position.
 - Optional. From the **Significance** list, select a significance.
15. To modify settings for a phrase, either double-click a phrase or select it and click the **Edit** button.

7 Classification

7.1 About Classification

In classification, you define classes to which WebCenter Forms Recognition assigns documents.

In classification, the confidence level indicates the degree of similarity between a document to be processed and the documents in the learnset of a given class.

WebCenter Forms Recognition can only assign a document to a class if a number of confidence values exceed predefined thresholds. Although you can influence the threshold values and the evaluation algorithms, the default settings are fine for most cases.

- The confidence values indicate the degree of similarity between the properties of a document and the properties of a class.
- A document can be classified if its confidence value with respect to a given class exceeds a predefined threshold. By default, this threshold is 70%.
- In general, there is only one target class per document. If several classes have a high confidence with respect to a document, it must be possible to distinguish reliably. Therefore, a certain difference in confidence between the winning class and the second-best competitor is required as well. By default, this *distance* is 20%.
- If multiple classification methods are applied to a given class, WebCenter Forms Recognition will compute a combined result. Ensure all documents to be processed within the same project have been scanned to the same resolution. If Brainware Table Extraction is to be used the resolution of all documents must be 300 DPI.

7.2 Advanced Classification

The *Classification Matrix* functionality can be configured in Designer for Verifier users. This feature will suggest several probable classes if the class result could not be determined by 100%. If more than one class is in the list, the class entry is supplemented by a value representing the determination probability. You will find the class with the highest probability on the top of the list.

7.2.1 Enable Advanced Classification

You can make this functionality available in Verifier by checking **Enable Advanced Classification in Verifier** on the **Verifier Mode** tab in the project's *Settings* dialog. To access the *Settings* dialog, use the **Options** menu and select **Settings...**

7.3 Prepare Sample and Test Documents

Before you begin designing your application, you need to ensure that you have enough documents to setup your project and to test it. To prepare the sample and test documents, complete the following steps:

1. Create a directory for sample documents.

2. Create a directory for test documents.
3. Within the sample directory, create subdirectories for each class.
4. Copy the documents to the corresponding subdirectories.
5. Create separate batches for each subdirectory by performing only the Import and OCR process steps.

7.4 Create a Project

To set up a new project, complete the following steps:

1. Start WebCenter Forms Recognition Designer.
2. Set the project options as defined in the following list.
 - Make sure that manually trained documents can be added to the learnset.
 - Specify or create a *Definition Mode* or *Runtime Mode* export directory. The software needs this directory to save processing results.
 - Specify or create a *Train Mode* base directory. The software will write copies of your sample documents to this directory.
 - Specify the batch root directory. The software will take the document input from there. The batch root directory name must not contain multi-byte characters.
 - Set all other options as desired.
3. Save the project file.

7.5 Create Classes

WebCenter Forms Recognition Designer enables you to create multi-level hierarchical classification schemes by means of base classes and derived classes. Base classes constitute the highest level of a classification tree. Derived classes are specializations of base classes.

When you create a class hierarchy, consider that:

- Create derived classes should be a true specialization of their parent class.
- Disjoint classes that compete with each other.
- Classes that compete with each other should be based on a similar level of abstraction.
- A working classification tree must contain at least two classes.
- Do not reuse the name of a deleted, learned document class in the same project. Doing so may cause inconsistencies within the project. When you create a document class, WebCenter Forms Recognition creates a directory with the name of the document class that contains the learnset documents for this class. This directory remains when you delete the class from the class hierarchy. If you create a new class with this name, it appears as if learnset documents are available when they are not.

7.5.1 Create a Classification Tree

To create a classification tree, complete the following steps:

1. On the toolbar, click the **Switch to Definition Mode** button.

2. To add a base class to your project, right-click any entry on the **Classes** tab. If your project does not yet contain classes, the only available entry is the root node, either labeled as **<new project>** or with the project file name.
3. On the popup menu, select the **Insert Base DocClass** option.
4. In the **Create New Base DocClass** dialog box, enter a class name.

Note: A valid name consists of alphanumeric characters without spaces or special characters. Later, you can define a different display name in the document class properties. For more information refer to the [Section 7.5.2: User-Friendly Display Names](#).

5. Click **OK** to confirm. The new class is inserted below the root class in alphabetical order.
6. Repeat steps 1 - 4 until all required base classes are created.
7. To add a derived class, right-click the prospective parent class. The parent may be either a base class or a derived class itself.
8. On the popup menu, select the **Insert Derived DocClass** option.
9. Enter a class name and confirm. The new class is inserted below the parent class in alphabetical order.
10. Repeat the steps 6 - 8 until all required derived classes are created.

7.5.2 User-Friendly Display Names

Display names enable you to specify user-friendly class names as a replacement for project class names.

This feature is important for some Supervised Learning configurations and usage of the Learnset Manager tool, in case the reviewed system-generated class names do not have meaningful names.

Both the Verifier and Designer applications use this display names in the list view items and the status bar indicator:

- The **Fields** view in *Definition Mode* of the Designer application, including both status bar indicators of the document class of the opened document it is currently classified to.
- In *Normal Train Mode* and *Verifier Test Mode*, the status bar indicator at the right of the document class of the opened document it is currently classified to.
- Document verification mode of Verifier, the status bar indicator at the right of the document class of the currently validated document it is classified to.
- The *Manual Classification* view of the Verifier application, including the list of the available classes and the currently selected class.
- *Global Learnset Browsing Mode* of the Learnset Manager tool.

Note: The other class view modes remain unaffected.

7.5.3 Edit Classes

WebCenter Forms Recognition Designer provides commands for editing classes. Exercise caution when using them since you might lose some of the settings you have already made.

The prerequisites for this task are:

- Designer is in *Definition Mode*.
- The panel on the left side of the window displays the **Classes** tab in the foreground.
- There must be classes to edit.

The following operations are supported:

- The **Delete DocClass** command from the shortcut menu of a class deletes the class including all associated settings. In this case, you lose all trained knowledge for the project. Repeat the training.
- The **Rename DocClass** command from the shortcut menu of a class renames the class.
In this case, you lose an existing learnset for the class and all trained knowledge for the project. Create a new learnset for the class and repeat the training.
- Within a classification tree, you can **Move** classes using drag-and-drop. In this case, you lose all trained knowledge for the project. Repeat the training. You may also have used inherited field properties. Check whether your settings are still consistent and meaningful.

7.6 Classification Methods

Once you have created classes, specify the classification methods for each class. Try a single classification method first. *Brainware* is the recommended primary method. If you need to improve the results, you can use complementary methods. Classification methods can be combined freely.

7.7 Select Classification Methods

The prerequisites for this task are:

- Designer is in *Definition Mode*.
- The panel on the left side of the window displays the **Classes** tab in the foreground.
- On the right side of the window, the *Classification* tab is selected.

To select one or more classification methods, complete the following steps:

1. On the **Classes** tab on the left side of the window, select a class.
2. In the **Classification** tab, on the right side of the window, check the methods that you want to use for the selected class. Clear the corresponding check mark to exclude a previously selected classification method.
3. Repeat the previous steps until all classes have classification methods assigned.

7.8 Project-Level Standard Classification

Standard classification uses the threshold and distance parameters defined at the project level to decide whether a document can be classified.

- Threshold defines the minimum confidence level.
- Distance defines the minimum difference in confidence between the best class and the second-best competitor.

Both requirements must meet to classify a document.

WebCenter Forms Recognition supports three ways to do this:

- [The Maximum method.](#)
- [The Average method.](#)
- [The Weighted Distance method.](#)

7.8.1 The Maximum Method

In this method, only the maximum confidence level of a class is taken into account. This is the default setting.

	Method A	Method B	Result
Class A	88%	68%	88%
Class B	67%	60%	67%
Class C	62%	63%	63%

Assume a document is classified with the following result:

- Class A: 88%
- Class B: 67%
- Class C: 63%

Now assume that the default values for threshold and distance are applied:

- Threshold: 70
- Distance: 20

Class A is the best class. Confidence for *Class A* is above the threshold.

Class B is the second-best class. The distance to *Class A* is **21**. This is above the required distance.

The document will be assigned to *Class A*.

7.8.2 The Average Method

The average (mean) of all confidence levels for a class is taken into account.

	Method A	Method B	Result
Class A	88%	68%	78%
Class B	67%	60%	63.5%
Class C	62%	63%	62.5%

Assume another document is classified with the following result:

- Class A: 78%
- Class B: 63.5%
- Class C: 62.5%

Now assume that the default values for threshold and distance are applied:

- Threshold: 70
- Distance: 20

Again, *Class A* is the best class. Confidence for *Class A* is above the threshold.

Class B is the second-best class, but for this document the distance to *Class A* is **14.5**. This is below the required distance.

This document cannot be classified using standard classification.

7.8.3 The Weighted Distance Method

The maximum of all results for all classes is taken as a reference. For each result, the distance to this reference is calculated. For each class, the distances are added. This sum is subtracted from the maximum result obtained for that class.

	Method A	Method B	Result
Class A	88 (0)	68 (20)	$68 = 88 - (0+20)$
Class B	67 (21)	60 (28)	$18 = 67 - (21+28)$
Class C	62 (26)	63 (25)	$8 = 63 - (26+25)$

In general, the percentage of documents that will be classified is as follows:

Maximum > Average > Weighted Distance

In particular, with the Weighted Distance method, it is highly unlikely that documents are assigned to a class if the class does not get high confidence levels from both methods applied

Assume document is classified with the following result, but this time the Weighted Distance method has been applied:

- Class A: 68%
- Class B: 18%
- Class C: 8%

Now assume that the default values for threshold and distance are applied:

- Threshold: 70
- Distance: 20

Class A is the best class but confidence for *Class A* is below the threshold.

Distances to the next classes are very high.

The document cannot be classified using standard classification

If you use the weighted distance method, you should work with smaller thresholds to obtain a satisfactory fraction of classified documents.

7.8.4 Project-Level Parent Classification and How It Works

So far, no hierarchical elements were taken into account for classification. However, if classification is not possible using the flat mechanism, hierarchical elements may resolve the conflict.

One of the available hierarchical methods is *parent classification*. If parent and child classes compete for the same documents, the documents will rather be assigned to the children. If there are several children with similar confidence levels, the parents can decide which class is the correct one.

Parent classification uses two additional values defined at the project level:

- A parent threshold (50 percent by default).
- A parent distance (10 percent by default).

7.8.5 Project-Level Default Classes and How They Work

Normally, all documents that cannot be assigned to a particular class are marked as **Not Classified** in Designer's *Runtime Mode*.

During production, they are marked by the Runtime Server with a particular status value indicating that classification was not successful. The corresponding documents will not be processed any further.

To prevent this, you can define a default class for each project that will be assumed if classification fails. You can also use this feature to create applications without classification, but with extraction.

7.8.6 Modify Project-Level Settings

The prerequisites for this task are:

- Designer is in *Definition Mode*.
- On the **Classes** tab on the left side of the window, the project node is selected.
- On the right side of the window, the tabs with class and field properties are visible.

On the **Classification** tab, the following options are available:

- To set a default class, select an entry from the **Default Classification Result** drop-down.

On the **Project** tab, the following options are available:

- To change the algorithm to obtain combined results for given classes, selects an entry from the **Classification Interpretation** list box.
- To change **Threshold** or **Distance for Standard Classification**, use the sliders or type values into the corresponding text boxes.
- To change **Threshold** or **Distance for Parent Classification**, use the sliders or type values into the corresponding text boxes.

7.9 Class-Level Sub-Tree Classification and How It Works

Subtree classification is a means to push classification from parent to child classes by defining reduced requirements for classification into the sub-tree.

Sub-tree classification uses two additional values defined at the class level:

- A sub-tree threshold (60 percent by default).
- A sub-tree distance (10 percent by default).

It is also possible to entirely exclude the parent class as valid classification result.

7.9.1 Class-Level Redirection and How It Works

Redirection means that in case of classification to a *Class A*, the documents will in fact be assigned to a different class such as *Class B*. Use redirection to combine several classes or in cascading classification schemes.

7.9.2 Modify Class-Level Settings

To change the class-level settings, complete the following steps:

1. Switch to *Definition Mode*.
2. On the **Classes** tab on the left side of the window, select a class.
3. On the right side of the window, select the **Classification** tab.
4. To assign the class to a view, click the **Brainware Classify Engine** method name
5. In the **Brainware Classify Engine** tab, select a view from the list box.
6. On the right side of the window, select the **Document Class** tab.
7. To use a different display name for the class, enter it into the corresponding text box.
8. To exclude the class as classification result, clear the **This DocClass Is a Valid Classification Result** check box.
9. To hide the class in Verifier, clear the **Visible in Correction** check box.
10. To push classification to the children of this class, check the **Force Sub-Tree Classification** check box. To change **Threshold** or **Distance** for sub-tree classification, use the sliders or type values into the corresponding text boxes.
11. To redirect classification results, from the **Redirection Classification Result** list, select a class.
12. Select the **Redirect after completion of extraction** check box.
13. To apply a static page count to all documents of this class, check the **Static Page Count** check box and specify the number of pages using the spin control.
14. Under **Table Training**, select either **Automatically using Learnset** or **Manually**.
15. From the **Classification Field** list, select the field which is used as classification field in Supervised Learning.
16. To setup validation rules, select the Validation tab.

7.10 Stretching

Stretching improves the results of *Brainware Classification* when combined with other classification engines. *Stretching* means to increase the distance between results given by the classification engine.

You can change the **Stretching** factor with the slide bar. The stretching factor, which increases the distance, can vary from **1.0** to **5.0**. For example, if the stretching factor is **2.0**, the classification result is as shown in the following table:

Class ID	Percentage Before Stretching	Percentage After Stretching
Class1	70%	70%
Class2	66%	62%
Class3	56%	46%
Class4	32%	8%

7.11 Calibration

Calibration fine-tunes the various classification engines. Calibration is based on the *Calibration factor*, which can vary from **0.1** to **1.5**.

7.12 Classification Learnset

The set of sample documents you need to provide for classification is called the *Classification Learnset*. Carefully select the samples for each class, because the quality of your learnset is crucial for the success of the classification.

When you select the sample documents for the learnsets, consider the following:

- You need a separate learnset for each class.
- You need samples that truly represent the topic you want to cover.
- Avoid documents containing multiple topics, since these documents cannot be classified exclusively to one class.
- In general, do not use the same document in more than one learnset.
- Avoid using similar documents for different classes.
- Use only documents with good OCR results for the learnset. When in doubt, highlight the OCR results to review the documents. Never use handwritten documents.
- You need at least five sample documents per class. The maximum number of samples per class is limited by memory.
- The number of samples needed to create a good learnset depends on the classification task. The broader a field a class is supposed to cover; the more samples are required. For example, you will need more samples to cover the topic *politics* than you will need for invoices of a company.
- Increase the number of samples if classes are very similar to each other.
- Ideally, you should have an equal number of samples for each class.

7.12.1 Classification Engine that requires Learnset

Following are the classification engines that requires learnset:

- Brainware Classify Engine
- Layout Classify Engine
- ASSA Classify Engine
- Brainware Layout Classification

7.12.2 Create a Learnset

The prerequisite for this task is “Add trained documents to the learnset” option is active on the *Train Mode* tab in the *Settings* dialog box.

To create a learnset, complete the following steps:

1. Switch to *Document Selection Mode*.
2. Select the first document from the batch that contains the documents for your learnset.
3. Switch to *Train Mode*. On the left side of the window, the **Classes** tab must be in the foreground.
4. The list in the lower section of the **Classes** tab indicates the number of documents in the learnsets.
5. If required, use the navigation buttons in the toolbar to browse to a document that you want to add to a learnset.
6. On the **Classes** tab, select a target class from the list box in the upper section.
7. In the toolbar, click the **Add to Learnset** button to add the current document to the classification learnset of the selected class. Designer writes a copy of the document to a protected directory structure that is reserved for learnsets. The entry in the **Classification Learnset** column in the list in the lower section of the **Classes** tab indicates that a document has been added.
8. Repeat steps 5 - 7 until you have assigned enough samples from the current document set.
9. To add more documents from other document sets, complete this entire process.
10. Save the project. This saves all changes to the learnset and updates the project.

7.12.3 Learnset Encryption

Use the learnset encryption for secure distribution of learnsets that contain confidential data.

You can activate the encryption of learnsets for any Verifier project.

You can use an encrypted learnset to train a project, or classify and extract documents, however you cannot display the workdoc files or the associated documents.

When you encrypt learnset, WebCenter Forms Recognition deletes all corresponding image files.

The result of the encryption process affects only the content of the learnset directories, where all workdoc files are going to be encrypted, as well as all the image

files – removed from the disk. Make sure to create a backup of the learnset before applying the encryption process.

Before the encryption process is applied learnset directory of a document class includes original image files and the workdoc document in a normal non-encrypted form.

When the encryption has been applied, learnset directories do not contain any images and the internal workdoc structure cannot be seen any longer.

You can use the feature for secure encryption of any learnsets, but it is supposed to be applied to so-called generic learnsets. When used for vendor learnsets, take into account that the class names are not encrypted and remain as they are.

After encrypting a learnset, you can still add documents to the learnset. You can also repeat the encryption step to encrypt the new documents.

7.12.4 Learnset Encryption Restrictions

Before you encrypt a learnset, consider the following restrictions:

- You cannot undo the encryption.
- Do not re-learn learnsets which use image based classification engines.
- Do not encrypt a learnset if you use image based classification engines. Re-learning a learnset deletes the workdoc files for the encrypted images and causes a failure.
- You cannot reapply the OCR process on an encrypted learnset, because the original image data is unavailable.

7.12.5 Encrypt Learnset

The prerequisite for this task is that Designer is in *Definition Mode* or *Train Mode* and backup of the learnset directory.

To encrypt a learnset:

1. From the **Options** menu, select **Encrypt Learnset**. A warning message appears.
2. Click **Yes** to encrypt the learnset.
3. In the **Training Manager** dialog, click **OK**.
4. After encrypting a learnset, you can still add documents to it. Repeat the previous steps to encrypt the new documents.

7.12.6 Custom Encryption Keys

If required, you can apply a custom encryption key in the Windows registry that the encryption process uses when encrypting the learnset documents. To access encrypted workdocs, use the *SCBCdrWorkdoc* method *EnableAccess* in a script.

If the custom encryption key is not available, the system uses an internal default key.

7.12.7 Apply Custom Encryption Key

To create the register key, complete the following steps:

1. Launch the Windows registry editor.
7. Complete one of the following substeps.

- For a 32-bit machine, navigate to [HKEY_LOCAL_MACHINE\SOFTWARE\Oracle\Cedar].
 - For a 64-bit machine, navigate to [HKEY_LOCAL_MACHINE\SOFTWARE\Wow6432Node\Oracle\Cedar].
8. In the right pane, right-click and then click **New > String Value**.
 9. In the Name field, type **EncryptLearnSetKey** and then click **OK**.
 10. Right-click the **EncryptLearnSetKey** key and then click **Modify**.
 11. In the **Edit String Value** dialog box, in the **Value data** field, type the custom encryption key and then click **OK**.

Note: The maximum character length for the key is 255.

7.12.8 Learnsets Correction

Occasionally, a document is added to a learnset by mistake, or it is assigned to the wrong class.

To review the learnset:

1. Select **Learnset** as document input in *Document Input Selection Mode*. A list view of classes is displayed.
2. Double-click one of the classes to display the documents in its learnset in a second list view below.
3. Using the list view of documents, you can now:
 - Browse through the documents. Click one of the documents to display it.
 - Delete a document from the learnset. Right-click the document and select **Remove from Learnset** from the popup menu.
 - Move a document to the learnset of a different class using drag-and-drop.
 - Exclude a document from the learning for specified classification methods. Clear the check mark in the corresponding column. This method is also useful for temporarily removing a document from the learnset.

7.12.8.1 Manually Add New Documents to a Learnset

Although you can add new documents to a learnset at any time, you may find that the results are inconsistent when you add new documents to a learnset and then save the project without learning. This only happens when you use a project that has a history of learning, as the classification engine has a knowledge base from which to extract information.

Without proper learning, the information from the new documents will not be included in the classification. To remedy this, you can change a learnset by manually adding a document to it and then changing the working mode. The documents will automatically go through the learning process once the mode changes.

Use this method for small projects and demonstration purposes only. For large projects in which learning is a lengthy task, refer to [Section 5.1.5: About Learning](#).

To add a document to a learnset manually, complete the following steps:

1. Switch to *Document Input Selection Mode* and select a document from the batch.
2. Switch to *Train Mode*. On the left side of the window, the **Classes** tab must be in the foreground.

The list in the lower section of the **Classes** tab indicates the number of documents in the learnsets. If required, use the navigation bar in the toolbar to browse to a document to be added to a learnset.

3. On the **Classes** tab, select a target class from the list box in the upper section.
4. In the toolbar, click the **Add documents to Learnset** button to add the current document to the classification learnset of the selected class. WebCenter Forms Recognition creates a copy of the document to a protected directory structure that is reserved for learnsets. The entry in the *Classification Learnset* column in the list in the lower section of the **Classes** tab indicates that a document has been added.
5. Repeat the previous steps until you finish adding all preferred documents.
6. Select a working mode. The system automatically learns the new documents.
7. Save the project.

7.12.9 About Learning

During Learning, WebCenter Forms Recognition uses your learnsets and the manually assigned target classes to identify the correlation between document input and target class. WebCenter Forms Recognition learns to classify documents automatically if they are similar to the documents in one of the learnsets.

Learning is required:

- After you have set up your classification scheme and created learnsets.
- When you have changed the classification tree by adding, deleting, moving or renaming classes.
- When you have added or removed a classification method that relies on learning.
- When you have changed any parameters affecting these classification methods.
- When you have changed a learnset.

7.12.10 Use the Learn Mode

The prerequisites for this task are:

- You must have set up a classification scheme with the minimum number of classes.
- You must have specified classification methods that rely on learning.
- You must have created appropriate learnsets with the minimum number of samples.

To use the learn mode, complete the following steps:

1. Switch to *Train Mode* or *Definition Mode*.
2. If required, select the **Options** menu and then **Incremental Learning** to change the scope of learning. If this option is enabled, learning only covers changes to the learnset. If this option is disabled, the entire project will be learned.
3. In the toolbar, click the **Learn Documents** button.

Note: Learning may take several minutes.

7.12.11 Check the Learn Status of a Class

Learning always attempts to cover the entire classification tree. Only classes with insufficient learnsets are omitted. You can check at any time whether a class has already been learned or whether it requires re-learning.

To check the learn status of a class, complete the following steps:

1. Switch to *Definition Mode*.
2. In the left pane, on the **Classes** tab, right-click a class and then click **Show Properties**.
3. In the **Classification** tab on the right side of the window, under **Used Classification Engines for this DocClass**, select an engine.
4. If the engine state in the lower right pane displays **Class is currently not trained**, click the Learn documents button.

7.13 Brainware Layout Classification Engine

The Brainware Classify Engine provides the ability to distinguish documents by their content.

You can use the Brainware Classify Engine as the only classification engine.

The formal structure of the document is unimportant. The same is true for syntax or phrases, since only words are used as input.

You can solve the following classification problems with the Brainware Classify Engine.

- Emails classified by project
- Business correspondence by type, such as orders, invoices, or resumes
- Customer requests classified by topic

7.13.1 Configure Brainware Classify Engine

To configure the Brainware Classify Engine, complete the following steps:

1. Switch to *Definition Mode*.
2. In the left pane, on the **Classes** tab, right-click a class and then click **Show Properties**.
3. In the right pane, on the **Classification** tab, under **Used Classification Engines for this DocClass**, select **Brainware Classify Engine**.
4. In the left pane, on the **Classes** tab, select the project.
5. In the right pane, on the **Classification** tab, under **Used Classification Engines for this project**, select **Brainware Classify Engine**.
6. From the **Brainware View** list, select a view or click **Insert** to insert a new view.
7. If required, modify the **YES Threshold** value to define the minimum confidence for a class within the view to return **Yes**.

8. If required, to return **Yes**, modify the **YES Distance** value to define the minimum distance to the second best class within the view.
9. If required, to return **No**, modify the **NO Threshold** value to define the maximum confidence for a class within the view.
10. Under **Word Statistic**, specify parameters for words that you want to include in the view's dictionary that embodies the base for learning with WebCenter Forms Recognition. In general, you do not have to change these values.
 - **Min. word length** Minimum number of characters in words. Shorter words will not be taken into account.
 - **Min. word frequency** Minimum number of occurrences in the learnsets of this view. Words that occur less frequently will not be considered.
11. Optional. To restrict the classification to certain areas within the documents, modify the settings on the **Regions** tab.

Note: For more information, refer to [Section 7.13.2: Restrict Dictionary Creation](#).

12. Optional. On the **Tuning** tab, modify the values for **Stretching** and **Calibration**.

7.13.2 Restrict Dictionary Creation

You can restrict the text base for dictionary creation to certain areas within the document. This may speed up processing as no full-page OCR is carried out. To use this advantage, make sure that the Runtime Server settings do not override your restrictions.

For more information on *OCR settings in Runtime Server*, refer to *WebCenter Forms Recognition Runtime Server Guide*.

To restrict the dictionary creation to certain areas within the documents, complete the following steps:

1. On the Regions tab, select the **Restrict engine to region** check box.
2. Select the pages that are affected by the restriction.
3. For each selected page, enter the regions that are to be taken into account:
 - The entire area is specified by Top 0 percent, Bottom 100 percent, Left 0 percent, Right 100 percent.
 - Top 20 percent, bottom 100 percent covers the bottom 80 percent of the document, cutting 20 percent off from the top.
 - Top 0 percent, Bottom 80 percent, covers the top 80 percent of the document, cutting 20 percent off from the bottom.
 - Top 100 percent, Bottom 100 percent, Left 100 percent and Right 100 percent equals an empty area.

7.13.3 Create a View for the Brainware Classify Engine

For instructions on how to assign classes to views, refer to [Section 7.9.2: Modify Class-Level Settings](#).

Each WebCenter Forms Recognition project automatically contains a view called *Default View*. By default, all classes use this view.

To create a view, complete the following steps:

1. Switch to *Definition Mode*.
2. On the **Classes** tab, select the project.
3. The **Classification** tab displays on the right side of the window.
4. Under **Used Classification Engines for this project list**, select **Brainware Classify Engine**.
5. On the **Brainware Classify Engine** tab, click **Insert**.
6. The *New Brainware View* dialog displays, enter a name without spaces for the new view.
7. Click **OK**.

7.14 Layout Classify Engine

The Layout Classify Engine classifies documents using the mosaic technique.

Use this engine before the OCR processing to pre-classify and rule out documents that do not match.

7.14.1 Configure Layout Classify Engine

To configure the Layout Classify Engine, complete the following steps:

1. Switch to *Definition Mode*.
2. In the left pane, on the **Classes** tab, right-click a class and then click **Show Properties**.
3. In the right pane, on the **Classification** tab, select the **Layout Classify Engine**.
4. In the left pane, on the **Classes** tab, select the project.
5. In the right pane, on the **Classification** tab, select the **Layout Classify Engine**.
6. Optional. On the Image **Classify** tab, modify the **Threshold** and **Distance** parameters.
7. Optional. On the **Restricted Area** tab, to restrict the classification to certain areas within the documents, complete the following sub-steps:
 - a. Select **Restrict engine to region**.
 - b. Specify the regions you want to take into account.
8. Optional. On the **Algorithm** tab, specify the required resolution.

7.15 ASSA Classify Engine

Like the *Brainware Classify Engine*, the *ASSA Classify Engine* needs a learnset of documents for classification. The difference between the *Brainware Classify Engine* and the *ASSA Classify Engine* lies in the different search pools that are generated out of the documents of the learnset when the documents are learned.

For its search pool, the *Brainware Classify Engine* generates a dictionary of words. The documents for the search pool are coded with words from this dictionary.

The *ASSA Classify Engine* uses the *trigram* method to code the search pool. The classification of documents is based on the block text search, which involves retrieving the text from the document block-by-block and generating a query string out of the block text. The *ASSA Classify Engine* then searches for that query within the search pool.

Configure the *ASSA Classify Engine*, complete the following steps:

1. Switch to *Definition Mode*.
2. In the left pane, on the **Classes** tab, right-click a class and then click **Show Properties**.
3. In the right pane, on the **Classification** tab, select the **ASSA Classify Engine**.
4. In the left pane, on the **Classes** tab, select the project.
5. In the right pane, on the **Classification** tab, select the **ASSA Classify Engine**.
6. Optional. On the **ASSA Classification** tab, modify the **Threshold** and **Distance** parameters.
7. Optional. On the **Regions** tab, to restrict the classification to certain areas within the documents, complete the following sub-steps:
 - a. Select **Restrict engine to region**.
 - b. Select the required pages.
 - c. For each selected page, specify the percentage that you want to take into account.
8. Optional. On the **Tuning** tab, use the sliders to modify the settings for **Stretching**, **Calibration**, and the **Quality/Speed** relation.

7.15.1 Quality / Speed

This option allows the project administrator to select a rapid classification, which sacrifices quality, or a slower classification, which offers greater accuracy. To elect rapid classification, you can set a percentage of the full block text. Or you can elect greater accuracy by creating a query based on the full block text, which takes additional time to compute.

7.16 Brainware Layout Classification Engine

Use the Brainware Layout Classification Engine for content or type classification. The Brainware Layout Classification (BLC) Engine provides a more precise classification between documents with similar templates. For example, considering invoices from different vendors it would be possible to reach a better result by taking into account the positional information of the documents' content as well as the textual content of the documents.

The document is divided into a number of zones. Each piece of text is tagged to indicate in which zone it appears. This means that if for the learned layout class X the word "VAT" was mostly located in the top left-hand corner region and for the another class Y it is usually in the bottom right-hand corner, the system would prefer one

classification over the other while for normal Brainware classification such a difference between X and Y would have been irrelevant.

The Brainware Layout Classification engine requires more sample documents in the learnset than the Template Classification engine. To achieve best classification weightings, it is important to have a sufficient number of documents in the learnset.

Setting up Brainware Layout Classification is similar to setting up Brainware Classification. You just need to provide sample documents for learning.

If there is already a learnset for Brainware Classification, there is usually no need to create another one for Brainware Layout Classification.

To create a separate learnset for Brainware Layout Classification, proceed as described in [Section 5.7.1: Create a Learnset](#). To carry out the learning, proceed as described in [Section 5.7.5: Learning](#).

The Brainware Layout Classification engine is an engine for the purpose of so-called *layout* classification using the powerful Brainware Classifier technology. While the Brainware Classification engine available in WebCenter Forms Recognition is used for *content* or *type* classification, the purpose of Brainware Layout Classification engine is to provide more precise classification between documents with similar templates, like, for example, invoices delivered from different vendors, where it is possible to reach more exact result by taking into account the positional information of the documents' content and not only the textual content of the documents like in ASSA or Brainware *content* classification engines.

The Brainware Layout Classification engine simply applies normal Brainware Classification. However, each word of both learned and extracted documents is not just used as-is, but extended with special character sequences that uniquely identify which zone of the document the word belongs to. Visually this idea can be represented as a document split to a number of zones. In the illustration below, the yellow area is the first page of the document and **VAT** is the word we are looking for.

	A ₁	B ₁	C ₁
A ₂		VAT	
B ₂			
C ₂			

Now the **VAT** word is going to be extended as **VAT_BBBB_AAAA**, identifying that we look for the word **VAT** only in the {B₁; A₂} region of the document. This then simply means that if for the learned layout class 'A' the word **VAT** was mostly located in region {C₁; C₂} and for other class 'B' it was usually {B₁; A₂} the system will prefer class 'B' while for normal Brainware Classification such a difference between the two classes would have been irrelevant. Exactly this very simple approach allows to the content classification engine (in this case Brainware Classifier) as more precise layout one.

7.16.1 Configure Brainware Layout Classification Engine

To configure the *Brainware Layout Classification Engine*, complete the following steps:

1. Switch to *Definition Mode*.
2. In the left pane, on the **Classes** tab, right-click a class and then click **Show Properties**.

3. In the right pane, on the **Classification** tab, select **Brainware Layout Classification Engine**.
4. In the left pane, on the **Classes** tab, select the project.
5. In the right pane, on the **Classification** tab, select **Brainware Layout Classification Engine**.
6. Optional. On the **Brainware Layout Classification** tab, from the **Brainware View** list, select a view or click **Insert** to insert a new view
7. If required, to return **Yes**, modify the **YES Threshold** value to define the minimum confidence for a class within the view.
8. If required, to return **Yes**, modify the **YES Distance** value to define the minimum distance to the second-best class within the view.
9. If required, to return **No**, modify the **NO Threshold** value to define the maximum confidence for a class within the view.
10. Optional. To restrict the classification to certain areas within the documents, modify the settings on the **Regions** tab.

Note: For more information, refer to [Section 7.13.2: Restrict Dictionary Creation](#).

11. Optional. On the **Extended** tab, modify the settings.

7.16.2 Extend Brainware Layout Classification Specific Settings

To change the extended settings for the *Brainware Layout Classification* engine, complete the following steps:

1. Click the **Extended** tab.
2. Under **Extended Settings**, set the following values:

Horizontal Covering Frequency	The number of zones each document is divided into across the page.
Vertical Covering Frequency	The number of zones each document is divided into down the page.
Do Not Differentiate Between Digits	Treat 123 and 987 as the same sequence of digits of the same length. Use this option when most numbers do not have specific meaning, for example, in the tables of a typical invoice document. Leave this option unchecked if the project's documents contain specific sequences of digits that can be critical for layout classification e.g., bank account information specific to each vendor class.
Attach Delimiters to Words	Check this option, in most cases, delimiters are essential to distinguish between words. An American format date, for example, may be better treated as a single differentiating word (00/00/000) in the classification's surrounding compared with 3 <i>noise</i> words 00 , 00 and 0000 .

Note: Save and relearn the project to apply new settings.

7.16.3 Create a View for the Brainware Layout Classification

For instructions on how to assign classes to views, refer to [Section 7.9.2: Modify the Class-Level Settings](#). Each project automatically contains a view called *DefaultView*. Without user intervention, all classes use this view. To create a view, complete the following steps:

1. Switch to *Definition Mode*.
2. In the left pane, on the **Classes** tab, select the project.
3. In the right pane, on the **Classification** tab, from the **Used Classification Engines for this project** list, select **Brainware Layout Classification**.
4. On the **Brainware Layout Classification** tab, click **Insert**.
5. In the **New Brainware View** dialog box, type a name without spaces for the view and then click **OK**.

7.17 Forms Classification

To use *Forms Classification*, you must create reading zones on documents and compare the recognition results from these zones with known identifiers of a class of forms. You can also configure a *Forms Classification* on the project level if more than one class shares zones.

Reading zones are fixed areas on documents that contain information that is to be recognized. In WebCenter Forms Recognition, three different types of reading zones exist:

- OCR
- Barcodes
- OMR

However, for *Forms Classification*, only OCR and barcode zones can be used.

For **OCR** zones, the contents of the reading zone are determined using optical character recognition. The textual content of the zone is the result of this process.

For **Barcode** zones, the contents of the reading zone are determined using barcode recognition. The number of strings represented by the barcode is the result of this process.

To learn about OMR zones, refer [Section 10.8.1: OMR Zones](#).

7.17.1 Configure Forms Classify Engine

To configure the Forms Classify Engine, complete the following steps:

1. Switch to *Definition Mode*.
2. In the left pane, on the **Classes** tab, right-click a class and then click **Show Properties**.
3. In the right pane, on the **Classification** tab, select the **Forms Classify Engine**.
4. In the left pane, on the **Classes** tab, select the project.

5. In the right pane, on the **Classification** tab, select the **Forms Classify Engine**.
6. On the **Forms Classify** tab, select the interpretation type.






7.17.2 Create Reading Zones

The prerequisites for this task is:

- You must have set up a classification scheme.

You may use more than one reading zone on each form. To create reading zones, complete the following steps:

1. Load a document into the viewer that belongs to the selected class.
2. Switch to *Definition Mode*.
3. In the left pane, on the **Classes** tab, right-click a class and then click **Show Properties**.
4. In the right pane, on the **Classification** tab, select the **Forms Classify Engine**.
5. On the **Classes** tab on the left side of the window, select a class.
6. From the menu, select **View** ▾ **Show Page**. The viewer toolbar displays additional buttons for zone creation.

Button	Description
	Activates the selection tool.
	Creates an OCR reading zone.
	Creates an OMR reading zone.
	Creates a barcode reading zone.
	Creates an anchor.

7. In the viewer panel, click the relevant toolbar button to create an OCR zone or a barcode zone.
8. Click the document and drag to create a rectangular reading zone around the form identifier. Make the rectangle large enough to hold the identifier. It should not be much larger, though. The reading zone displays as a transparent rectangle with red borders. Each zone has a label that displays the name of the zone. The initial name is created from the word **Zone** and a consecutive number.
9. In the left pane, select the project.
10. In the right pane, on the **Classification** tab, select the **Forms Classify Engine**.
11. On the **Forms Classify** tab, select one of the following options.
 - **Binary classification.** Possible results of forms classification are Yes and Maybe.
 - **Weighted classification.** Forms classification returns a confidence value as result.

7.17.3 Configure Form Identifiers

To configure the form identifiers, complete the following steps. These form identifiers are applicable only for *OCR zones*:

1. Switch to *Definition Mode*.
2. In the left pane, on the **Classes** tab, right-click a class and then click **Show Properties**.
3. In the right pane, on the **Classification** tab, select the **Forms Classify Engine**.
4. On the **Forms Classify** tab, complete the following sub-steps:
 - a. Click a cell of the **Zone Name** column and select a zone from the list of available reading zones.
 - b. In the **Compare Text** column, type the string that identifies your form. This string is compared to the recognition result from the reading zone.
 - c. In the **Page Number** column, click to specify whether the reading zone is on the first or on the last page of the document.
5. Repeat previous steps for all reading zones that should be considered.

7.17.4 Specify Compare Method

By default, the recognition result and form identifier are compared using an error-tolerant algorithm.

To specify the comparison method in more detail, complete the following steps:

1. On the **Forms Classify** tab, click the row number to select a row.
2. Click **Advanced**.
3. In the **Compare Text** field, you can edit the comparison string. This may be required, depending on the comparison method you select.
4. In the **Compare Method** list, select one of the methods. For more information, refer to [Section 7.17.5: Compare Methods](#).
5. You may want to assign several forms with similar identifiers or several versions of the same form to the same class. In this case, it may be useful to ignore leading or trailing characters of the recognition string. You can set the number of characters that are to be neglected using the **Ignore First** and **Ignore Last** spin controls.
6. The **Max. Compare Distance** allows you to perform the comparison, neglecting leading or trailing characters of the comparison string. The compare distance is computed as follows:

$$\text{Compare Distance} = \mathbf{1.00} - \text{Number of Characters (Recognition String)} / \text{Number of Characters (Comparison String)}$$






Note: A match requires that the actual compare distance is less than or equal to the maximum compare distance.

7.17.5 Compare Methods

String Compare	A very simple method that returns a match for a literal occurrence of the comparison string.
Trigram	An error-tolerant method that returns a match for a literal occurrence of the comparison string, but also for strings that can be derived from the specified one by fragmenting the text into groups of three characters called trigrams. The number of identical groups determines whether there is still a match.
Levenshtein	Another error-tolerant method that returns a match for a literal occurrence of the comparison string, but also for strings that can be derived from the specified one by inserting, interchanging or deleting single characters. This is the default method.
Regular Expression	A method that returns a match if the recognition result corresponds to a format pattern specified in the Compare Text field.
Simple Expression	A method that returns a match if the recognition result corresponds to a simple format pattern specified in the Compare Text field.

7.17.6 Color-coding in the Compare Text Column

On the **Forms Classify** tab, color-coding is applied to the comparison string to indicate the selected method:

Color	Method
 Purple	String Compare.
 Turquoise	Trigram.
 Olive	Levenshtein.
 Gray	Regular Expression.
 Black	Simple Expression.

7.17.7 Global OCR Zones in Forms Classification

It is possible to share OCR zones from the project page between different document classes. If there are several document classes that have to read the forms ID from the same place in the document, you can configure the project to read the zone globally.

Before you share OCR zones, make sure to have at least one class set up with *Forms Analysis*.

7.17.7.1 Setup a Shared OCR Zone

A shared OCR zone defined on project-level appears in the Zone Name list for all classes that have forms classification. To set up a shared OCR zone, complete the following steps:

1. Select the project node in the **Classes** tree.
2. Decrease the size of the **All** zone to a small rectangle.
3. Set up a shared OCR zone by completing the steps in [Section 7.17.2: Create Reading Zones](#).

7.18 Phrase Classification

The Phrase Classify Engine uses words and phrases to identify the document class. Within certain error limits, the phrases or words must literally occur within the documents.

This engine is a very powerful complement to content classification since you can define certain keywords that must lead to a specific class. You do not have to define every possible phrase, since the engine does that job.

You can use this engine as the only classification engine. However, this requires the set up of and maintenance of a potentially complex set of rules. Typically you use the Phrase Classify Engine in conjunction with the Brainware Classify Engine.

Note: Do not use the Phrase Classify Engine to train base classes.

7.18.1 Configure Phrase Classification

To configure the Phrase Classify Engine, complete the following steps:

1. Switch to *Definition Mode*.
2. In the left pane, on the **Classes** tab, right-click a class and then click **Show Properties**.
3. In the right pane, on the **Classification** tab, select the **Phrase Classify Engine**.
4. In the left pane, on the **Classes** tab, select the project.
5. In the right pane, on the **Classification** tab, select the **Phrase Classify Engine**.
6. Optional. On the **Phrase Classify** tab, modify the **Threshold** and **Distance** parameters.
7. Optional. On the **Regions** tab, to restrict the classification to certain areas within the documents, complete the following sub-steps:
 - a. Select Restrict engine to region.
 - b. Select the required pages.
 - c. For each selected page, specify the percentage that you want to take into account.

7.18.2 Search Phrase

A *search phrase* consists of one or more *search words* (text strings). When a *search phrase* consists of multiple words, separated by spaces, it is called a *search expression*. As such, a space character cannot be used for any other purpose and cannot be searched for explicitly.

The following characters are special wildcard characters:

*	Represents zero or more characters within a single word.
?	Represents a single character (but not a space).
??	When preceded and followed by spaces, represents zero or one word. If ?? is included within a <i>search word</i> , it will represent two characters.

Similarly, these wildcard characters cannot be searched for explicitly.

Leading or trailing wildcard characters on the search word are not meaningful and will be removed or ignored. For example, the search word *read* will match any word that contains *read*, such as *bread*, *read*, *reader*, or *breeding*. Search words that include leading or trailing wildcard characters, such as *?read* or *read** will not result in any additional matches.

Note: The wildcard word, **??**, can be entered manually, or can be configured through the Designer user interface by setting a *Maximum distance to next word*. This value is the distance between this *search word* and the next. For example, a value of **1** means the *search words* must be in adjacent words, and a value of **6** would allow for up to 5 words to appear between this and the next *search word*.

7.18.3 Candidates

A candidate is a set of possible values for a field.

When a *search phrase* consists of a single *search word*, any word in a document that contains the *search word* as a substring is identified as a candidate. The candidates for the selected *search phrase* will be highlighted on the document.

For each *search word*, an allowable number of errors (up to 8, depending upon the length of the search word) may be defined. An error may be a missing character, an inserted character, or a substituted character. The number of errors for a candidate is indicated in the tooltip for that highlighted candidate on the document.

It can be sometimes confusing how errors are applied. For example, if the search word is **qntity** and one error is allowed, the word **quantity** will be picked up as a candidate, because **qntity** matches substring **antity** with one error.

When searching for a *search expression*, the same method is applied to each search word ensuring that the *Maximum distance to next word* between search words is not exceeded.

Note: For search expressions with three or more search words, intermediate search words can be made optional through the **Word Significance** slider.

If the **Significance** of a search phrase is set to *forbidden*, and at least one reliable candidate is found, candidates will not be highlighted.

7.18.4 Phrase Weight

For each candidate, a weight is calculated based on the significance/strength settings for the *search word(s)*, as well as the frequency of errors, if any. The weight for a *search phrase* is the maximum weight of all of its candidates, scaled according to the **search phrase's** Significance setting.

Note: The number of candidates for any of the **search phrases** does not affect phrase weight, though it is taken into account when using Phrase Classify Engine for Classification.

This phrase weight can be seen in the tooltip of any one highlighted candidate in the document, generally a percentage between **1** and **100**. However, in the case of a 'forbidden' **search phrase** for which all candidates are questionable, such as contain errors, the phrase weight may be less than zero.

7.18.5 Classification Value

When the *Phrase Engine* is used in the *Classification* step, the output is based upon the candidate weights and/or phrase weights.

If the document contains a forbidden search phrase, the classification is set to **No** for this engine. Otherwise, the engine returns a *Classification* value, generally between **0** and **100**.

In all cases, weights are scaled by a factor of **100** to arrive at a *Classification* value. When the sum is greater than **100**, the *Classification* value is set to **100**.

7.18.6 Add a Phrase

To use phrase classification, you must add keywords or phrases that lead to a particular class or exclude a document from a class.

The prerequisites for this task is:

- You must have set up a classification scheme.

To add a phrase to the phrase classification, complete the following steps:

1. Switch to *Definition Mode*.
2. In the left pane, on the **Classes** tab, right-click a class and then click **Show Properties**.
3. In the right pane, on the **Classification** tab, select the **Phrase Classify Engine**.
4. Optional. To add a new phrase manually, on the **Phrase Classify** tab, in the **New Phrase** field, type a keyword or a phrase and then click **Ins**.
5. Optional. To add a word or a phrase from the document, complete the following sub-steps:
 - a. On the toolbar, click the **Highlight Words** button.
 - b. In the document, click a word to copy it into the **New Phrase** field.
 - c. Optional. To build a phrase, click another word.
 - d. Click **Ins**.
6. Optional. To restrict the analysis area, in the **Position** column, select one of the following options:

Note: Restricting the analysis area may improve the performance, as no full-page OCR takes place. Ensure that the Runtime Server settings do not override your restrictions. The phrase is found if any part of the phrase extends into the selected half.

- To restrict the analysis area to the top half of the page, select **Top**.
 - To restrict the analysis area to the bottom half of the page, select **Bottom**.
 - To restrict the analysis area to the left half of the page, select **Left**.
 - To restrict the analysis area to the right half of the page, select **Right**.
7. Optional. To highlight the occurrences of a phrase in the active document, complete the following sub-steps:
- a. On the toolbar, click the **Highlight nothing** button.
 - b. Select the phrase from the list of active phrases.

Note: Each occurrence is highlighted in green.

- c. Optional. To view the number of typing errors and the phrase weight, point to the highlighted phrase. On the **Classes** tab on the left side of the window, select a class.

7.18.7 Edit a Phrase

To edit a phrase, complete the following steps:

1. Switch to *Definition Mode*.
2. In the left pane, on the **Classes** tab, right-click a class and then click **Show Properties**.
3. In the right pane, on the **Classification** tab, complete the following sub-steps:
 - a. Select a phrase from the list of active phrases and then click **Edit**.
 - b. In the **Phrase Settings** dialog box, modify the settings. The dialog box provides a separate tab for each word in the phrase.

Note: For information about the phrase settings, refer to [Section 7.18.8: Phrase Settings](#).




7.18.8 Phrase Settings

The settings you can adjust are:

Tolerance	Maximum number of typing errors allowed for this word. If the phrase contains the word invoice , for example, and the number of error characters has been specified at 2 for this word, the words invoke , involve and invite would also meet the requirements.
Word Strength	The stronger the word, the more typing errors can occur until the word loses its meaning. A maximum strength stands for a word that retains full significance with the maximum number of permissible typing errors. In general, you should keep the default setting for the word strength.

Accept Only Case Sensitive Matching	Select this option to enforce case-sensitivity.
Relative Significance	Use this option to set the proportional significance of the current word within the phrase. For example, if your phrase is with kind regards , you could reduce the relative significance of the word with . As you move the associated slider, the weight will be shown to the right in plain text.
Maximum Distance	Use this setting to specify whether your phrase will still be recognized if there are additional words between the current one and the next one specified. For example, you could use a single definition for please pay by and please pay this amount by , with a maximum distance of 3 . For consecutive words, the distance is 1 .
Significance	This setting affects the entire phrase. The significance may vary between absolute (i.e., a document belongs for sure in this class if the phrase is found) and forbidden (i.e., a document cannot belong in this class if the phrase is found). Be careful when using the extreme values. Use them only on very strong phrases and try to avoid them on single words.
Function	Specifies the function that applies if a phrase is found more than once in the same document. This setting affects the entire phrase. To change the function, click the button.

The method used to calculate the *Classification* value depends upon the function selected for each *search phrase*.

Button		Description
	Degrassive	The sum of all candidate weights. Note: For any search phrase with one or more candidate, the effective weight of each subsequent candidate is reduced proportionally, e.g. a second candidate will count on half, a third candidate will count one third.
	Linear	The sum of all candidate weights. The full significance of the phrase is scored with every further occurrence.
	Binary	The sum of all phrase weights. A multiple occurrence is not scored higher than a single occurrence. Note: If multiple candidates are available in the search area, the method works with only one candidate, which is in most cases the lowest on the image.

7.18.9 Test a Phrase

To test a phrase, complete the following steps:

Prerequisite: Configure the learnset manager base directory in the project settings to enable the Test and Test All buttons.

1. Switch to *Definition Mode*.
2. In the left pane, on the **Classes** tab, right-click a class and then click **Show Properties**.
3. In the right pane, on the **Classification** tab, complete one of the following sub-steps:
 - Select a phrase from the list of active phrases and then click Test. The Phrase test dialog box displays the occurrences of the selected phrase for all learnset documents for all classes.
 - Click **Test All**. The **Phrase test** dialog box displays the occurrences of all phrases for the learnset documents of the selected class.

7.18.10 Delete a Phrase

To delete a phrase from the phrase classification, complete the following steps:

1. Switch to *Definition Mode*.
2. In the left pane, on the **Classes** tab, right-click a class and then click **Show Properties**.
3. In the right pane, on the **Classification** tab, select a phrase from the list of active phrases and then click **Del**.

7.18.11 Export Phrases

You can export phrases to import them in another project. To export phrases to a file, complete the following steps.

1. Switch to *Definition Mode*.
2. In the left pane, on the **Classes** tab, right-click a class and then click **Show Properties**.
3. In the right pane, on the **Classification** tab, select **Export**.
4. In the **Save As** dialog box, in the **File name** field, type a file name and then click **Save**.

7.18.12 Import Phrases

To import phrases from a file, complete the following steps.

1. Switch to *Definition Mode*.
2. In the left pane, on the **Classes** tab, right-click a class and then click **Show Properties**.
3. In the right pane, on the **Classification** tab, select **Import**.
4. In the **Open** dialog box, select an EXP file and then click **Open**.

7.19 Language Classify Engine

The Language Classify Engine automates the classification of processed documents by language.

The engine contains predefined and pre-trained learnsets for the following languages:

- Danish
- English
- Finnish
- French
- German
- Greek
- Russian
- Spanish
- Swedish

The English, French, German, Greek and Russian languages are specifically optimized for invoice processing.

You can add new languages and adjust the predefined ones.

The Language Classify Engine uses the ASSA content classification technology to distinguish between processed documents written in different languages. The engine imports the data in Unicode format and converts them to non-western language learnsets with a special phonetic English representation suitable for the core ASSA engine. After this conversion, the engine creates an ASSA classification pool that is then used for classification purposes. The pre-trained version of this pool is distributed along with **WebCenter Forms Recognition** setup and is available without having to learn anything, although it is adjustable if the customer would like to extend the predefined learnset.

When classifying the next incoming document, the system will use the available ASSA classification pool for the classification. During the classification, the engine ignores all numeric words found in the document since they are usually irrelevant in terms of language classification.

To use the engine for the non-western languages, complete the steps in Activate multi-byte encoding for CJKT languages.

7.19.1 Configure Language Classify Engine

To configure the *Language Classify Engine*, complete the following steps:

1. Switch to *Definition Mode*.
2. In the left pane, on the **Classes** tab, right-click a class and then click **Show Properties**.
3. In the right pane, on the **Classification** tab, select the **Language Classify Engine**.
4. In the left pane, on the **Classes** tab, select the project.
5. In the right pane, on the **Classification** tab, select the **Language Classify Engine**.
6. Under **Languages available for this project**, review the selected languages.
7. To set the learnset location, complete one of the following sub-steps:
 - **Application Path** Stores language classification learnset in *[Installation Folder]\Bin\bin\Langpool* and the language source fields in *[Installation Folder]\Bin\bin\Langpool\Source*.

- **Project Learnset Path.** Stores the language classification learnset and source files in the project learnset directory, as configured on the **Train Mode** tab in the **Settings** dialog box.
8. If you selected **Project learnset path**, in the project learnset directory, and manually create the following directories.
 - \LangPool
 - \LangPool\Source
 9. Ensure that the \LangPool\Source directory includes all necessary LNG files. (see Help ▾ Classification ▾ About the Language Classify Engine ▾ About new languages in the standard learnset)

Note: For more information, refer to [Section 7.19.3: About New Languages in the Standard Learnset](#).

10. To regenerate the entire ASSA classification learnset, click the **Create New Learnset** button.

Note: The *threshold* and *distance* settings used while evaluating the classification are taken from the **Standard Classification** settings defined on the **Project** tab of the project node settings.

1. **Project Learnset Path.** Stores the language classification learnset and source files in the project learnset directory, as configured on the **Train Mode** tab in the **Settings** dialog box.

Languages Available for This Project	Use this option to select the languages that are used in your project.
---	--

Learnset Location

Create New Learnset	Click this button to regenerate the entire ASSA classification learnset. For example, you can use this option to update the language classification configuration with new languages to be used for classification.
----------------------------	---

Note: The *threshold* and *distance* settings used while evaluating the classification are taken from the **Standard Classification** settings defined on the **Project** tab of the project node settings.

7.19.2 Create New Language Assignment for a Class

During classification, the system searches for class language assignments to determine the appropriate class. To create a new language assignment, complete the following steps:

1. Switch to *Definition Mode*.
2. In the left pane, on the **Classes** tab, select a class.

3. In the right pane, on the **Classification** tab, under **Used Classification Engines for this DocClass**, select **Language Classify Engine**.
4. Under **Languages available for this DocClass**, select one or more languages.

7.19.3 About New Language in the Standard Learnset

When generating a new learnset, the engine automatically searches for all available LNG files in the `[Installation Folder]\Bin\bin\Langpool\Source` directory.

During the generation of the pool, the newly imported languages get their names from the corresponding LNG files.

7.19.3.1 Create a Unicode Language Input File

To create new Unicode input source language file, complete the following steps:

1. Launch **Microsoft Word**.
2. Open the language text file.
3. On the **File** menu, click **Save As**.
4. In the **Save As** dialog box, from the **Save as type** list, select **Plain Text** and then click **Save**.
5. In the **File Conversion** dialog box, under Text encoding, select **Other encoding**.
6. From the encoding list, select **Unicode** and click **OK**.
7. Change the file extension to LNG.
8. Copy the LNG file to the `[Installation Folder]\Bin\bin\Langpool\Source` directory.

7.19.3.2 Add a New Language to the Learnset

To add a new language to the learnset, complete the following steps:

1. Switch to *Definition Mode*.
2. In the left pane, on the **Classes** tab, right-click the project and then click **Show Properties**.
3. In the right pane, on the **Classification** tab, select the **Language Classify Engine**.
4. Select the newly added language.
5. Click **Create New Learnset**.

7.19.3.3 Exclude a Language from the Learnset

To exclude the language from the list of possible classification results, complete the following steps:

1. Switch to *Definition Mode*.
2. In the left pane, on the **Classes** tab, right-click the project and then click **Show Properties**.
3. In the right pane, on the **Classification** tab, clear the Language check box.

4. Optional. To remove the language from the ASSA classification learnset, click **Create New Learnset**.

7.19.3.4 Remove a Language from the Learnset

To remove a language from the language classification learnset you can perform one of the following options:

1. In the *[Installation Folder]\Bin\bin\Langpool\Source* directory, delete the corresponding LNG file.
2. Switch to *Definition Mode*.
3. In the left pane, on the **Classes** tab, right-click the project and then click **Show Properties**.
4. In the right pane, on the **Classification** tab, clear the Language check box.
5. Optional. To remove the language from the ASSA classification learnset, click **Create New Learnset**.

7.19.3.5 Modify an existing Language Learnset

To modify the learnset for an existing language, complete the following steps:

6. Open the LNG file in Microsoft Word.
7. Modify the file.
8. Save the file as Unicode file.
9. Switch to *Definition Mode*.
10. In the left pane, on the **Classes** tab, right-click the project and then click **Show Properties**.
11. In the right pane, on the **Classification** tab, click **Create New Learnset**.
12. **New Learnset**.

7.19.4 Check Classification Results

WebCenter Forms Recognition uses the standard classification rules when applying the classification process with the Language Classify Engine.

To check the exact classification results and weights on a per language basis, complete the following steps:

1. In the left pane, on the **Classes** tab, select a language class.
2. On the toolbar, click the **Show/hide classification matrix** button.
3. In the **Class Result Matrix** dialog box, verify the results and then click **Close**.

7.20 Image Size Classification Engine

The Image Size Classification engine uses the physical size of documents to distinguish classes. This method is not able to prove positively that a document is in a specific class. Instead, image size classification rules out classes and whole branches by excluding documents that do not match. This method is very fast. Use it to reduce a classification problem and to shorten the computing time.

7.20.1 Configure Image Size Classification

To use image size specification, you must specify the document size for the corresponding class. It is also possible to specify a size range.

The prerequisites for this task is that you must have set up a classification scheme.

To configure the Image Size Classification engine, complete the following steps:

1. Switch to *Definition Mode*.
2. In the left pane, on the **Classes** tab, right-click the project and then click **Show Properties**.
3. In the right pane, on the **Classification** tab, select **Image Size Classification**.
4. In the **Image Size** tab, from the **Minimum Size** list, select a size.
5. From the **Maximum Size**, select a size.
6. Under **Orientation**, select the page orientation.
7. Under **Tolerance**, specify the error tolerance that engine uses when comparing sizes.

Note: This value is particularly important if your **Maximum Size** and **Minimum Size** are equal.

7.21 Template Classify Engine

The Brainware Layout Classification engine has replaced the Template Classify Engine. For more information, refer to [Section 7.13: About the Brainware Layout Classification Engine](#).

7.22 Engine-Level Absolute Results

For a certain classes and documents, each classification method may return different result types. Possible result types are:

Confidence	This document belongs to a specified class with a probability that is given in percent.
Yes	This document is definitely in this class.
Maybe	This document might belong to this class.
No	This document does definitely not belong to this class.

The various classification methods are not equally reliable. For example, it is dangerous to decide that a document is an invoice based only on the fact that it is a certain size. The table below presents the possible results for each classification method. It also shows which result type is returned by default.

Method	Result	Default or Configured
Brainware Classify	Confidence	Default
	Yes	Configured

Method	Result	Default or Configured
	No	Configured
ASSA Classify	Confidence	Default
	Yes	Configured
	No	Configured
Phrase Classify	Confidence	Default
	Yes	Configured
	No	Configured
Image Size Classification	Maybe	Default
	No	Default
Form Classify	Yes	Default
	Maybe	Default
	Confidence	Configured
Layout Classify	Confidence	Default
	Yes	Configured
	No	Configured
Brainware Layout Classification	Maybe	Configured
	No	Configured
	Confidence	Default
Language Classify	Maybe	Configured
	No	Configured
	Confidence	Default

Yes and **No** are called *absolute results*. If the program is configured accordingly, they are obtained if confidence levels returned by one of the methods are either very high or very low.

If a method returns **Yes** for a class, the combined result will be **Yes**. Classification of the document will stop at this stage. All other classes will be excluded automatically.

If a method returns **No** for a class, the combined result will be **No**. Within the current branch, classification of the document will stop at this stage. All child classes will be excluded automatically.

You can use absolute results to speed up the entire classification process. *Image Size Classification* is a suitable method to exclude entire branches very quickly. *Brainware Layout Classification* can be used to positively confirm formal classes very quickly. The same is true for *Forms Classification*.

7.23 Classification Testing

After you set up a classification scheme and define and configure the classification methods, you can test the classification.


7.23.1 Test the classification in Definition Mode and Train Mode

The prerequisites for this task are:

- There is an active document set.

- There is a trained classification scheme.
- To test the classification only, your project settings should exclude data extraction from runtime functionality.

To test the classification and to determine target classes and confidence values use one of the following buttons:

Button	Description
	Processes the current document. The button's drop-down menu enables or disables the debug mode.
	To Process and export the current document, from Classify/analyze the current document list, select Execute & Export .
	To Process and OCR the current document, from the Classify/analyze the current document list, select OCR & Execute .
	Processes the next document.
	Processes documents in the current set starting with the current one.

7.23.2 Debug the classification in Definition Mode and Train Mode

If you start the classification in debug mode, a dialog box with a tree view of possible target classes and confidence levels displays for each processed document. To debug the classification, complete the following steps:

1. From the **Classify / analyze the current document** list, select **Debug**.
2. In the **Debug Classification** dialog box, monitor the processing.
3. To proceed to the next document, click **Continue**.
4. Optional. To finish the debug mode, clear **Show this dialog** and then click **Continue**.

7.23.3 Classification results

The bottom left pane displays one of the following classification results for the current document:



1. For unprocessed documents, the software displays **Not Classified**.
2. If the classification could not assign the document: **Uncertain Classification**.
3. For successfully classified documents: **The proposed class**.


7.23.4 Display Classification Matrix

To display an array of classification results for the previously processed document, click the **Show/Hide Classification Results Matrix** toolbar button.

7.23.4.1 Test Classification in Runtime Mode

Use one of the following buttons to determine target classes and confidence values:

Button	Description
	Processes the next document.
	Processes documents in the current set starting with the current one.

	Stops the processing of documents. Clears any results from previous runs.
---	---

7.23.4.2 Classification Results in Runtime Mode

The results are presented in a panel that displays summary results for all classes within the current project. It contains two separate tabs, one for a statistic view of average (mean) classification and extraction results, the other one for a tree view of the classification distribution. Both views are continuously updated at runtime.

The classification-related columns on the **Statistic View** tab have the following meaning:

Column	Description
Class Hierarchy	Class name and position within the classification scheme.
Count	Number of documents assigned to a class, or total number of processed documents. Numbers are accumulated until processing is stopped.
Classification Level	Average (mean) confidence for assignments to a class.
Classification Difference	Comparison of actual classification results with results from a reference file. The first number represents the number of different classifications in this run; the second one represents the total number of classifications in this run. If there is a difference between test results and reference results, this is indicated by the red color.
Classif Level	Actual number of classifications vs. maximum number of classifications for the selected class in percent.
Classification Diff.	Class-level comparison of actual extraction results with results from a reference file. The first number represents the number of differences; the second one represents the total number of extractions. If there is a difference between test results and reference results, this is indicated by the red color.

The extraction-related columns in the lower left pane on the **Statistic View** tab have the following meaning:

Column	Description
Field Recognition Rate	Actual number of extractions vs. maximum number of extractions for the selected document in percent.
Extraction Diff	Document-level comparison of actual extraction results with results from a reference file. The first number represents the number of differences; the second one represents the total number of extractions. If there is a difference between test results and reference results, this is indicated by the red color.

7.23.4.3 Create a Reference File

The **Statistic View** tab enables you to compare your classification results with results stored as reference files. To create a reference file, select one of the following options:

Situation	Steps
Enable creation of the default export file.	<ol style="list-style-type: none"> 1. On the Options menu, click Settings. 2. In the Settings dialog box, on the Runtime Mode tab, select Generate default export file for each batch and then click OK.
Create a reference file from the Statistic View tab	<ol style="list-style-type: none"> 1. On the Statistic View tab, right-click a class and then click Save As Reference. 2. In the Save As dialog box, in the File name field, type a file name

Situation	Steps
	and then click Save .

7.23.4.4 Modify a Reference File

You can edit reference files, as well as the default export file, with a text editor, or you can use text editors to create a reference files from scratch. Each line in the file is tab-separated and represents the results of one document. The line structure is as follows:

```
<file name><TAB><class><TAB><field 1> ... <field n>
```

7.23.4.5 Load a Reference File

To load a reference file, select one of the following options:

1. Right-click any class entry on the **Statistic View** tab and select **Load Reference** from the popup menu, then browse to and select a reference file or the default export file.
2. In the **Open Reference File** dialog box, select a REF file and then click **Open**.

7.23.4.6 Compare Classification Results

To compare the classification results with results stored as reference files, complete the following steps:

1. On the **Statistics View** tab, right-click a class and then click **Compare with Reference**.
2. In the **Open Reference File** dialog box, select a REF file and then click **Open**.
3. On the **Statistic View** tab, select **Compare with**.
4. On the toolbar, click the **Process documents starting with the current one** button.
5. Review the compare results in the lower right pane.

7.23.4.7 The Class Tree View tab

The **Class Tree View** tab displays the distribution of class assignments. To show or hide a branch, double-click the corresponding root node.

If you click a class in either the *Statistic View* or the *Class Tree View*, refer to the panel for information regarding which documents have been assigned to the current class. The classification-related columns in this panel are defined in the following table.

Column	Description
Classification Level	Confidence for assignments to a class.
Ref. Classification	Target class for this document as given in the reference file.

7.23.4.8 Display the Class Tree View tab

To display the Class Tree View tab, complete the following steps:

1. In the upper left pane, click the **Class Tree View** tab.
2. To hide a branch, double-click the corresponding root node.

3. To show a hidden branch, double-click again.

7.24 Classification Problems

Most classification problems are due to one of the following causes:

- Insufficient OCR quality.
- Insufficient learnset quality.
- Inappropriate selection of classification methods.

7.24.1 Resolve OCR Problems

Use the highlighting options to identify whether there are problems with the OCR. If this is the case, proceed as follows:

1. Improve the quality of the images, for example ensure to scan the document with a resolution of 300 DPI or review your scanner settings.
2. The default OCR settings may not be suitable to process your document input. For instructions on how to change these settings, refer to [Section 9.14: Specify OCR Settings](#).

Note: Changing the OCR settings at the project level is a severe intervention. Make sure that other reasons are excluded before you do this.

7.24.2 Resolve Learnset Problems

To resolve learnset problems, complete the following steps:

1. Check the contents of unclassified or wrongly classified documents and compare them with the documents in your learnset.
2. To improve the results, add unclassified documents to the learnset of the target class in compliance with the following rules:
 - Add documents with a confidence below the threshold if their content fits into the selected class.
 - Add documents that were previously not classified correctly due to the initially insufficient learnset.
 - Use documents with an almost identical confidence for two or more classes. These documents are suitable to differentiate the classes.
 - Avoid documents with a high confidence. These documents do not improve the learnset, because they would already be classified correctly.

7.24.3 Resolve Classification Problems

To review and improve the classification results of problematic documents. Complete the following steps:

1. Switch to *Definition Mode*.
2. Identify the method that causes the problem.
3. Determine whether classification fails because the confidence is too low.
4. If documents are classified to the wrong class, modify the learnset or the parameters for the method that returns the wrong result.

5. If too many documents are not classified at all, consider adding another classification method, or providing more samples for the learnset.
6. If this does not help either, you may have to change the settings for classification evaluation.

8 Recognition Techniques and Engines

In WebCenter Forms Recognition, you can use the following recognition techniques:

- OCR (Optical Character Recognition).
- Barcode Recognition.
- OMR (Optical Mark Recognition).

Recognition is performed on demand during the classification and extraction steps. On demand means recognition can be performed as requested, any number of times. The result is stored in the workdoc.

Each engine has its default parameter set that should work well in most cases. However, to enable you to select the optimum parameters for your problem, if this should be required, you can customize the recognition settings independently on several levels for each processing step that involves recognition.

OCR is applied to provide text input for the classification step. In the extraction step, either of the recognition techniques can be used.

Depending on the recognition method, WebCenter Forms Recognition provides several recognition engines you can select from:

- For OCR:
 - FineReader 11.
 - FineReader 10.
 - Kadmos5 OCR/ICR Engine.
- For OMR:
 - Cairo OMR Engine.
- For Barcode Recognition:
 - Cleqs Barcode Engine.

8.1 FineReader OCR Engines

WebCenter Forms Recognition features the OCR engines FineReader 10 and FineReader 11. They optimize documents with poor or marginal printer quality, such as documents printed with dot matrix or chain printers.

- Do not use different FineReader engine versions in one project. This might cause process failures.
- When processing one document with multiple zones, if OCR is applied for multiple zones within a single page, the FineReader engine license counter increments only once for this page.

8.1.1 About FineReader Languages

The character set that the FineReader engines can recognize depend on the languages in the used languages list.

For information about the available languages, refer to *OCR Engine Languages* in the *WebCenter Forms Recognition Installation Guide*.

If **Digits** is in the list of used languages, the engines can recognize the following characters:

«()+, - ./0123456789:=#\$([{^?€ %).]}°¼½¾»-< >

If **English** is in the list of used languages, the engines can recognize the following characters:

!"#\$%&'()*+,-./0123456789:;<=>?@ABCDEFGHIJKLMN OPQRSTUVWXYZ
[\]_abcdefghijklmnopqrstu vwxyz{|}~£¥\$©«¬®°±»–„•€™

8.1.2 General Tab Settings

Recognition Trade Off is a tradeoff between performance and OCR accuracy, and has the following options:

Most Accurate	If selected, recognition has less error rate but a higher processing time. This is the default mode.
Balanced	If selected, recognition is in between Most Accurate and Fast .
Fast	If selected, recognition is quicker but less accurate.

The following *Rotation* options are available:

Detect Orientation	Select this to automatically detect whether pages are in portrait or landscape orientation.
Deskew	Select this to automatically fix the alignment of badly scanned pages.

The following *Enhancement* options are available:

Remove Lines	Select this option to remove all internal lines, either horizontal or vertical.
Remove Texture	Select this option to remove background noise from the image for recognition.
Despeckle	Select this option to remove speckles from your documents. Speckles are made up of a group of black pixels surrounded by white pixels or vice versa.

8.1.3 Recognition Tab

The following *Recognition* options are available:

Automatic	Select this option to automatically detect how the documents have been printed. As a rule, use this option unless your document input is very homogeneous and from a dot-matrix printer or a typewriter.
------------------	--

Matrix	Select this option for material printed with dot-matrix printers.
Typewriter	Select this option for material created with typewriters.
OCR-A / OCR-B	Select this option when the documents use OCR-A and OCR-B fonts. This is typically where high OCR character recognition rates are needed. The fonts are optimized for machine character recognition. For example, they are used on passports and other security documents.
MICR E13 B	A character recognition technology to facilitate the procedure of instance check.
Receipt	Allows for the detection of text specific to sales receipts, invoices, etc. This mode brings improvement for recognition of low quality documents where the OCR process focuses on tidying up the noises around the actual text. This option is only available for FineReader 11.
Allow Superscript	Select this option to allow for letters, characters, or symbols that are written above, or above and to the right or left of, another character. For example, the ™ symbol.
Allow Subscript	Select this option to allow a character that is printed on a level lower than the rest of the characters on the line, for example, the 2 in the chemical formula H₂O .
Small Digit for Asian Enhancement Box	This box is only accessible if at least one of the Chinese, Japanese or Korean languages is selected. This setting is only enabled if at least one of these languages is added to the list of used languages on the Language tab, such as ChinesePRC+English. For more information, refer to Section 4.6: About Non-Western Languages .

8.1.4 Add a Language to a FineReader Engine

The FineReader engines process the recognition with the languages in the used languages list. To add a language to the used languages, complete the following steps:

1. On the **Language** tab, from the **Installed** list, select a language and then click the arrow button to move it to the **Used** list.
2. Optional. Repeat the previous step to add other languages only if you are processing multilingual documents.

Note: For more details on available languages, refer to *OCR Engine Languages* in the *WebCenter Forms Recognition Installation Guide*.

8.2 Kadmos 5 Engine

The *Kadmos 5* engine optimizes recognition of handwritten forms and material containing handwritten zones. It offers support and optimization for different languages.

WebCenter Forms Recognition uses the Kadmos 5 Client Edition. The client version can be used by one Runtime Server instance at a time on the same machine.

A Server Edition is available. The server version can be used by multiple Runtime Server instances running concurrently on the same machine.

8.2.1 General Tab

The following *Recognition* options are available:

Recognition	The classifier determines the range of valid characters. These characters display on the Characters tab and you can further restrict them there.
Hand Alphanumeric German, Europe, UK, US	Select for German, European, UK or US hand-written material containing alphanumeric characters.
CMC7, E13B, Farrington F7B Normfont	Select for material written with CMC7, E13B (US checks), F7B (ID cards) fonts.
OCRA, OCRB	Select for material written with dedicated OCR fonts.
Hand + Machine Print, Numeric	Select for material containing handwritten or machine-printed numbers and currency symbols, but no other alpha characters.
Hand + Machine Print, Europe, US Numeric	Select for European or US documents.
Machine, Alphanum, German, UK, US	Select for UK, US, and German machine-printed material containing alphanumeric characters.
Confidence Level	Use the slider to specify the threshold for accepting characters. Characters recognized below the required confidence level are termed as rejects, characters recognized above the confidence level are termed as valid. High confidence level causes more reject, but there will be fewer errors in the accepted results. The confidence level will also affect performance. Generally, it takes about 10 times longer to read with the lowest confidence level than with the highest.

Recognition Trade Off	Select the type of optimization assigned to the ICR process. The <i>Most accurate</i> parameter provides higher accuracy with lower speed.
------------------------------	--

Info	Displays the Kadmos5 engine version, the date build, and the edition type.
-------------	--

8.2.2 Characters Tab

Restrict the range of valid characters. The offered character set depends on the classifier that was selected before.

- In the upper section, the **Valid Characters** tab displays a button for each possible character of the selected classifier. A pressed button corresponds to a valid character.
- In the bottom section, there are check boxes that enable you to edit entire character groups in one step. If a check box is selected, the corresponding character group is valid.
- Use the **All** and **None** buttons to edit the entire character set in one step. By clicking **All**, all characters are marked as valid. By clicking **None**, all characters are marked as invalid.

8.2.3 Digits Tab

Define the digits that are eligible to be found on the document. Use the **All** and **None** buttons to select and deselect all digits.

8.2.4 Preprocessing Tab

The Kadmos preprocessing features are implemented for special usage only. In most cases, it is recommended to use the WebCenter Forms Recognition preprocessing features as described in [Section 10.2: Modify the Settings on the Preprocessing tab](#).

Edit the following *Noise Removal* settings:

Automatic	The Automatic setting for noise removal is based on an algorithm that will produce best results for many scenarios depending on the characters' position, size and segmentation. For checks (cheques) use the Automatic setting. Only adjust the settings manually if specific values are needed for well-known cases.
------------------	--

Relative [%], Absolute [pixel]	Select one of these options to reduce background noise on your documents manually. You need to select the Enabled check box and to specify the maximum size of spots that are removed, either using the pixel scale or as a fraction of the line thickness of the found characters. The size parameters are: <ul style="list-style-type: none">▪ Diameter: Maximum diameter of a spot.▪ Thickness: Maximum horizontal length of a spot.▪ Length: Maximum vertical length of a spot.
---------------------------------------	--

Object Removal	Use this extension to remove lines or other geometrical shapes that could interfere with the OCR from the documents. <ul style="list-style-type: none"> ▪ Lines, horizontal or vertical ▪ Boxes ▪ Combs
-----------------------	--

For specific documents, it is recommended to use Lines Manager or Despeckling as Preprocessing first.

8.2.5 Position Tab

To specify settings with respect to word and line surroundings, select the **Position** tab.

These options are available:

Character Segmentation	If the characters in the zone are within equal or proportional distance, select Equal Distance or Proportional respectively. The Automatic setting is used by default. If the characters do not touch, check the No Touching Char. check box. If ligatures are expected in the zone, deselect No Ligatures .
Character Distance	Speed up processing by specifying the horizontal character distance in quantity of image pixels. Keep in mind that the character distance is resolution dependent. To activate the option, check the Defined check box. Specify the minimum and the maximum distance in image pixels using the associated text field.
Character Position	Select one of the options to take the position of a character within a line into account. The scale goes from irrelevant (Ignore) up to very relevant (Strong Weight).

8.3 Cairo OMR Engine

8.3.1 OMR Zones






Using the *Cairo OMR Engine*, you can determine the contents of OMR reading zone using optical mark recognition. The purpose of this process is to detect whether a check box on a document is marked or not. This is done by analyzing the black to white ratio of the defined zones. Therefore, an OMR reading zone can contain only one check box. If you have more than one check box on the document, you have to create an OMR reading zone for each of them.

8.3.2 Create OMR Reading Zones

To create reading zones, complete the following steps:

1. Switch to *Definition Mode*.

2. Load a document into the viewer that belongs to the selected class.
3. From the menu, select **View ▾ Show Page**. The viewer toolbar displays additional buttons for zone creation.

Button	Description
	Activates the selection tool.
	Creates an OCR reading zone.
	Creates an OMR reading zone.
	Creates a barcode reading zone.
	Creates an anchor.

4. On the toolbar, click the **Insert OMR zone** button.
5. In the document, click and drag to create the reading zone around the check box.

8.3.3 Modify the Cairo OMR Engine Settings

You need to specify threshold values to distinguish unchecked from checked areas. If the engine cannot decide whether the OMR zone is checked, such as for all degrees of blackness outside the ranges specified, it returns a question mark, similar to the behavior of OCR engines with unrecognized characters. Ensure that the OMR zone is not smaller than the specified zone, and that there is only one rectangle in the specified zone. In particular, if your OMR zones are filled in manually, do not choose too high values for the upper limit of checked OMR zones. A very high degree of blackness may rather indicate that a manual entry has been corrected. If this can be safely excluded, you can also configure the OMR recognition to never return question marks. Perform the following steps:

1. Switch to *Definition Mode*.
2. On the **Classes** tab on the left panel of the window, select a class.
3. Select the **Fields** tab on the left panel of the window.
4. On the right side of the window, select the **Fields** tab.
5. Click the **OCR Settings...** button.
6. In the **Field OCR Settings Properties** dialog box, under **Recognition Type**, select **OMR**.
7. On the General tab you can see the settings for the *Cairo OMR* engine. The following options are available to determine unchecked and checked areas:

Lower [%]	Lower limit for degree of blackness in percent.
Upper [%]	Upper limit for degree of blackness in percent.
Return Value	A number or a character string. This value is returned if the measured degree of blackness is between the lower and the upper limit for <i>Unchecked</i> or <i>Checked</i> , as appropriate.

Search for OMR Box	An algorithm searches for a black square box within the defined zone. For better recognition results, the recognition is then applied for this square box zone. If nothing was found, the engine searches for a check mark within the originally defined zone. If an OMR box was found, the returned value corresponds to the specified Return value. The returned value is “?” if no OMR box was found.
---------------------------	--

8.4 QualitySoft Barcode Engine

The QualitySoft Barcode Engine is WebCenter Forms Recognition’s default barcode recognition engine. It includes the following features.

- It reads all common barcode types.
- It recognizes multiple barcode on a single page.
- It reads barcodes oriented at any angle.

8.4.1 General Tab

The following options are available on the General Tab:

Max. number of barcodes	The maximum number of barcodes to recognize within the reading zone. The default is 10. Reducing the number increases the recognition speed.
Min. number of chars	The barcode minimum length, in characters.
Max number of chars	The barcode maximum length, in characters.
Orientation	The barcode’s orientation angles. Parameter options are 0, 90, 180, and 270 degrees. Clearing specific orientation values causes WebCenter Forms Recognition to ignore barcodes with that angle during processing. Complement the Orientation settings with the max. Skew settings to get most fast and accurate results.
Intensive Search	This setting triggers an additional barcode-detection algorithm which improves the barcode detection, but reduces the processing speed. Use this setting if your document set includes very small barcodes or if the recognition failed during an initial processing without this option selected.
Max. Skew	These settings complement the Orientation settings. They define the maximum angle deviation from the configured main orientations. Setting a small value increases recognition speed.

8.4.2 Types Tab

It is possible to select multiple types. Excluding barcode types which your documents don't contain increases the recognition speed.

The engine supports the following linear barcodes.

Specific barcodes for barcode type "2 of 5 family"

- Datalogic 2 of 5
- Iata 2 of 5
- Industry 2 of 5
- Interleaved 2 of 5
- Inverted 2 of 5
- Matrix 2 of 5
- BCD Matrix

Specific barcodes for barcode type "EAN/UPC"

- EAN 8
- EAN 13
- UPC-A
- UPC-E

Specific barcodes for barcode type "Code 39 family"

- Code 32
- Code 39
- Code 39 extended

Specific barcodes for barcode type "Patchcode"

Specific barcodes for barcode type "2D"

- PDF 417
- Datamatrix
- Codablock

Specific barcodes for barcode type "Misc"

- Code 93
- Codabar
- Code 128
- EAN/UCC128

8.5 Cleqs Barcode Engine

The *Cleqs Barcode Engine* is a barcode engine compatible with WebCenter Forms Recognition.

8.5.1 General Tab

The following options are available:

Threshold	The minimum confidence level.
Min. / Max. Number of Characters	Enter the minimum or maximum length of the barcode in characters.
Return Only Decodable Barcodes	Check this option to accept barcodes only if all characters are recognized, i.e. no rejects.
Max. Number of Barcodes	Enter the maximum number of barcodes to be recognized within the reading zone. The default is 10 . Reducing the number speed up processing.

8.5.2 Restrictions Tab

The following options are available:

Code Type	Select whether you want to recognize barcodes or patch codes. A patch code is a pattern of horizontal black bars separated by spaces and typically placed near the leading edge of a paper document. Patch codes are used to separate documents.
Barcode	Select this option to recognize barcodes and then select the barcode types you expect to find. Do not check more options than required as this slows down processing and affects accuracy.
Orientation	Select the appropriate option to specify whether all barcodes or patch codes are oriented horizontally, vertically or both.
Ink Color	Select the appropriate option to detect white barcodes on a black background, black barcodes on a white background or both.

8.5.3 Preprocessing Tab

The following Image Enhancement options are available:

No Isolated Pixels	Select this option to automatically remove small speckles from the reading zone.
No Hor. Single Pixel Gaps	Select this option to automatically fill small gaps in horizontal lines. This option is recommend for processing barcodes with a corresponding orientation.
No Ver. Single Pixel Gaps	Select this option to automatically fill small gaps in vertical lines. This option is recommend processing barcodes with a corresponding orientation.

The following Filter options are available:

Erosion

The erosion filter thins down lines by setting any black pixel to white if it is located at a specified direction to at least one white pixel in the source image. To specify the direction, select from the following options:

- Horizontal.
- Vertical.
- Isotropic (all directions).

Dilation

The dilation filter thickens lines by setting any white pixel to black if it is located at a specified direction to at least one black pixel in the source image. To specify the direction, select from the following options:

- Horizontal.
 - Vertical.
 - Isotropic (all directions).
-

8.6 OCR Field-Level Digit Booster Engine

The *Digit Booster* boosts digit-to-digit for alphanumeric fields and alpha-to-digit for numeric fields. The digit-to-digit booster provides optimized digit recognition. The alpha-to-digit booster also repairs digits that were recognized as letters.

Boosted objects are *worktexts* for fields, candidates and table cells. The workdoc itself remains unchanged. Multipage documents are supported but the engine speed may drop for multipage documents with a lot of fields and candidates. Too high resolutions and too low resolutions lead to boost failure.

The *Format Analysis Engine* delivers candidates for the *Brainware Extraction Engine*, so that re-learning is not required for Booster migration. For digit-to-digit boosting, relearning is not useful, although it can be helpful for alpha-to-digit boosting.

8.6.1 Booster Evaluation by Script

For booster evaluation by script, use the `ptrCandidate.Worktext.GetCharAttributes` method.

```
ptrCandidate.Worktext.GetCharAttributes(...)Confidence = (Confidence - 101) / 100.00
```

Always supply level 3 logs when reporting Digit Booster related issues or defects.

Always apply statistical evaluation of quality separately for the two boosting methods (Alpha to digit/digit to digit).

8.6.2 Activate the Digit Booster Engine

To activate the digit booster engine, complete the following steps.

Switch to *Definition Mode*.

1. In the left pane, on the **Classes** tab, select a class.
2. In the left pane, on the **Fields** tab, select a field.
3. In the right pane, on the **Field** tab, select **Boost field**.

4. Optional. Select **Boost digits only**.

8.6.3 Digit Booster Engine Options

The **Boost Field** option enables the alpha-to-digit and digit-to-digit boosting. Use this feature to correct a wrong recognition of numbers as letters. It should be used on numeric fields only, such as *Amounts* or *PO Number*.

Examples:

- 0: OCR'd as letter 'O'
- 5: OCR'd as 's'
- 3: OCR'd as 8

The **Boost Digits Only** option enables digit-to-digit boosting. Use this feature to correct a wrong recognition of numbers as numbers. It should be used on alpha-numeric fields like *Invoice Number* or *Date*.

8.6.4 Enable Extra Boosting

By Default, the boost process takes place in the *Document_PostExtract* event by default. If extra boosting is needed between the *Analysis* and *Evaluation* processes of the *Extraction* a registry value can be added as follows:

1. Launch the Windows registry editor.
2. Complete one of the following substeps.
 - For a 32-bit machine, navigate to [HKEY_LOCAL_MACHINE\SOFTWARE\Oracle\Cedar].
 - For a 64-bit machine, navigate to [HKEY_LOCAL_MACHINE\SOFTWARE\Wow6432Node\Oracle\Cedar].
3. In the right pane, right-click and then click **New > DWORD (32-bit) Value**.
4. In the Name field, type **ExtraBoostingPriorToEvaluationEnabled** and then click **OK**.
5. Right-click the **ExtraBoostingPriorToEvaluationEnabled** key and then click **Modify**.
6. In the **Edit DWORD (32-bit) Value** dialog box, in the **Value data** field, to enable the feature, type **1** and then click **OK**.

8.6.5 About Boosting Tables

The **Boost Digits Only** option does not work for table fields. This limitation can be circumvented by using script. For table fields use the column-level property *pTable.IsColumnNumeric* at the very end of the *Document_PostExtract* event. Set the *isColumnNumeric* property to **False** if alpha-to-digit boosting is not allowed.

8.6.6 About the Field-Level Digit Booster Candidate Evaluation

The digit booster engine enables an extended candidate highlighting mode indicating the boost result reliability by colors.

- **Green:** *Boosted* and *boosted secure* - the engine converts the character. The tooltip shows the action results and behavior.
- **Yellow:** The result is unreliable - the engine does not convert the character. The user must approve the result manually.
- **Red:** The result is completely unreliable, the engine does not convert the character.

Note: The advanced highlighting is not available for table cell boosting.

9 Data Extraction

To set up the data extraction, complete the following procedures:

- Creating data fields.
- Specify the analysis engine to obtain candidates.
- Specify the evaluation engine to select the correct candidate.
- Configure the selected engines.
- Apply validation settings.
- Test and optimize the data extraction.

A candidate can only be assigned to a field if its weight exceeds a predefined threshold. You can influence the threshold values and the evaluation algorithms, but in most cases, the default settings should be appropriate.

- The weight indicates the degree of similarity between the properties of a candidate and the properties of user-selected candidates.
- Data can be extracted if the weight of the best candidate with respect to a given field exceeds a predefined threshold. By default, this threshold is **50** percent.
- In general, there is only one successful candidate per field and document. If several candidates have a high weight with respect to a field and document, it must be possible to distinguish reliably. Therefore, a certain difference in weight between the winning candidate and the second-best competitor is required as well. By default, this *distance* is **10** percent.

Extraction means that selected data from a document is automatically written to an extraction file. In general, classification is a precondition for extraction because the fields that need to be extracted are usually different for each class. If it is not necessary to distinguish different document classes and only extraction has to be performed, you can carry out a dummy classification with only one class that needs to be defined as the default. For more information, refer to [Section 12.4: Project Level Default Class](#).

For each class, identify the business processes that use the documents. Identify the data that is required by subsequent systems. Then define the set of fields that are to be filled per class.

9.1 About Threshold and Distance

You can define threshold and distance values to configure confidence levels for validation.

The engine uses the threshold value at the evaluation step to decide if the extraction is successful. If the weight of the best candidate is less than threshold/100, extraction fails and no value is extracted.

The engine uses the distance value at the validation step to decide the validity of an extracted value. If the difference between the best and the second best candidates is less than distance/100, the extracted value is considered invalid, and displays in red in Verifier.

9.2 Field Types

Fields hold the data extracted from the documents. You define them on the class level. Fields and their properties are inherited along branches in the classification tree. All classes derived from a parent use the parent's field definitions.

9.3 Text Fields

You can use text fields with the Associative Search Engine, the Check Analysis Engine, the Format Analysis Engine, and the Zone Analysis Engine.

9.4 Table fields

You can use tables with the Brainware Table Extraction and the Table Analysis Engine.

9.5 Create a Field

When you create new fields, the system inserts the newest fields below existing ones in the order in which you create them.

To create a field for a class, complete the following steps:

1. Switch to *Definition Mode*.
2. In the left pane, on the **Classes** tab, select the class.
3. In the left pane, on the **Fields** tab, right-click and then click **Insert Field Definition**.
4. In the **Add Field** dialog box, type a field name and then click **OK**.

Note: Valid names consist of alphanumeric characters without periods, spaces or special characters.

5. To create additional fields, repeat the previous steps.

9.6 Create a Table Field

To create a table field, complete the following steps:

1. Create a field.

Note: New fields are created as text fields.

2. In the left pane, on the **Fields** tab, right-click the field and then click **Table**.
3. In the right pane, on the Analysis tab, under **Available Analysis Engines**, select **Brainware Table Extraction**.
4. Configure the **Brainware Table Extraction** engine.

Note: For more information, refer to [Section 9.18: About the Brainware Table Extraction Engine](#).

9.7 Specify a Display Name for a Field

To specify a display name for a field that Verifier displays in the status bar instead of the original field name. The class inherits the display name to all its subclasses.

Complete the following steps:

1. Switch to *Definition Mode*.
2. In the left pane, on the **Fields** tab, select a field.
3. In the right pane, on the **Field** tab, in the **Display Name** field, type the display name for the field.
4. On the **File** menu, click **Save Project**.

9.8 Delete a field

To delete a field, complete the following steps:

Note: If you delete a field, you lose all trained knowledge for the data extraction.

1. Switch to *Definition Mode*.
2. In the left pane, on the **Fields** tab, right-click the field and then click **Delete Field Definition**.
3. In the **Delete Field** dialog box, click **Yes**.
4. Repeat the training.

9.9 Check the Analyzed Table Content

To check the analyzed table content, complete the following steps.

1. Switch to *Definition Mode*.
2. In the left pane, on the **Fields** tab, press Shift and the table icon together.

9.10 Rename a Field

To rename a field, complete the following steps.

Note: If you rename a field, you lose an existing extraction learnset for the field and all trained knowledge for the project.

1. Switch to *Definition Mode*.
2. In the left pane, on the **Fields** tab, right-click the field and then click **Rename Field Definition**.
3. In the **Rename Field** dialog box, type a new name and then click **OK**.
4. If required, create a new learnset for the field and repeat the training.

9.11 Specify the Analysis Engine

After you have created fields, specify an analysis engine for each field. To select an analysis engine for a field, complete the following steps:

1. Switch to *Definition Mode*.
2. In the left pane, on the **Fields** tab, select a field.
3. In the right pane, on the **Analysis** tab, from the **Available Analysis Engines** list, select an engine.
4. Repeat the previous steps for all fields.

9.12 Project-Level Recognition Settings

The project-level recognition settings are assigned to a special reading zone, the All zone. This zone represents the entire page. The All zone applies to all pages of the document. The system automatically synchronizes the project-level and page-level settings.

Note: Do not modify the properties of the All zone directly, since some of the options, such as the Anchor options, do not make sense when applied to the entire document set.

Project-level settings affect:

- The classification step.
- Extraction steps using *Format Analysis*.
- Extraction steps using *Zone Analysis*, provided that no different settings have been specified at the zone level.

Project-level settings affect every document processed with the project. They influence both quality and performance. Therefore, exercise caution when changing them.

Modify the project-level settings if your classification does not work properly due to OCR problems, or if you have OCR-related problems getting candidates from *Format Analysis*. You also should edit the project-level settings to change the default engines for each recognition type.

9.12.1 Modify the Project-Level Recognition Settings

Classification and extraction methods that rely on learning use a dictionary created from the words in your documents. To modify the project-level recognition settings, complete the following steps:

1. Switch to *Definition Mode*.
2. In the left pane, on the **Classes** tab, right-click the project and then click **Show Properties**.
3. In the right pane, on the **Project** tab, click the **OCR settings** button.
4. In the OCR Properties dialog box, modify the settings on the **Preprocessing** and the **Recognition** tab and then click **OK**.
5. To change the rules for word segmentation, add or remove characters in the **Word segmentation characters field**.
6. Complete the following substeps:
 - a. On the toolbar, click the **Classify / analyze current document** button.

- b. Click any of the highlighting buttons and carefully observe the results of your changes.
 - c. Repeat the previous steps with a couple of sample documents.
7. If learning was processed before, repeat the training.

9.13 Modify the Field-Level Recognition Settings

Field-level recognition settings affect the extraction steps. You can use them to read the field again after the best candidate has been selected. By this time, the format and the geometry of the field is well known. Therefore, the OCR quality can be improved significantly by using a different OCR engine or by imposing certain restrictions through classifiers.

To modify the field-level recognition settings, complete the following steps:

1. Switch to *Definition Mode*.
2. In the left pane, on the **Fields** tab, select a field.
3. In the right pane, select the **Field** tab.
4. If the current field is a derived field, and you do not want to inherit the OCR settings of the parent class, clear **Use derived OCR settings**.

Result: This enables the remaining OCR parameters.

5. Optional. To activate OCR optimization, select **Read field again**.
6. Optional. Click the **OCR settings** button and then, in the Field OCR Settings Properties dialog box, modify the settings and then click **OK**.
7. Optional. To activate the digit booster, select **Boost field**.

Note: For details, refer to [Section 8.6: About the OCR Field Level Digit Booster Engine](#).

8. To check the results of field boosting, refer to [Section 8.6.6: About the Field-Level Digit Booster Candidate Evaluation](#).

9.14 Specify OCR Settings

To specify the OCR settings, complete the following steps:

1. Switch to *Definition Mode*.
2. In the left pane, on the **Classes** tab, right-click the project and then click **Show Properties**.
3. In the right pane, on the **Project** tab, click the **OCR Settings** button.
4. In the **OCR Properties** dialog box, on the **Preprocessing** tab, under **Available Methods**, select any of the following methods and then click the arrow button to move it to the **Selected Methods** list.
 - To convert images to black-and-white images, select Binarisation.
 - Box and Comb Removal
 - Clean Border

- Despeckle. For more information, refer to [Section 9.15: Despeckling](#).
 - To revert white-on-black pages to black-on-white pages, select Invert.
 - Lines Manager
5. If you selected **Box and Comb Removal**, complete the following substeps:
 - a. In the **Selected Methods** list, select **Box and Comb Removal**.
 - b. On the **Box or Comb Removal** tab, select both **Box Removal and Automatic**.
 6. If you selected **Lines Manager**, complete the following substeps:
 - a. In the **Selected Methods** list, select **Lines Manager**.
 - b. On the **General** tab, select **Remove Horizontal Lines, Remove Vertical Lines, and Repair Characters**.
 7. On the **Recognition** tab, under **Recognition**, select **OCR**.
 8. From the **Available Engines** list, select a **FineReader** engine.
 9. On the **General** and **Recognition** tabs, review the settings based on your document quality.
 10. On the **Languages** tab, select **Digits** and **English**, and if required, add other languages.

9.15 Despeckling

This process, which removes dots and visual dust from your documents, sometimes detects decimal points and removes them. This, of course, can be undesirable. If your documents are generally light, do your first preprocessing on them without despeckling to test the results. If the quality is still poor, enable the Bespeckle option in the OCR settings and preprocess the documents again.

9.16 About Line Extraction

9.16.1 Description

The line extraction engines extract lines from the document and stores them into the workdoc as TextLines.

The selected line extraction engine is triggered after OCR recognition and detection of word objects. It is also executed by the script method `pWorkdoc.RebuildBasicObjects`.

The line extraction engine delivers the results as a collection of `ISCBCdrTextBlock` objects accessible through the workdoc interface within a script using the method `pWorkdoc.Textline(LineIndex)`.

The following line extraction engines are available:

- Brainware Line Extraction. This is the recommended and default line extraction engine for all newly created projects.
- Standard Line Extraction. This engine might be active for projects created with older versions and is available due to backwards compatibility reasons.
- Advanced Line Extraction.

The selected line extraction engine impacts the Format Analysis engine as it uses internal line objects for formats recognition.

9.16.2 Switch between the Line Extraction Methods

To switch between the line extraction methods, complete the following steps:

1. On the **Options** menu, click **Settings**.
2. In the **Settings** dialog box, on the **Processing** tab, under **Line extraction settings**, select one of the options.

9.17 Associative Search Engine

9.17.1 About Associative Search Engine

The *Associative Search Engine* (ASE) matches document text to records within a *search pool*.

In many cases, the search pool contains a list of contract partners, such as vendor names and their addresses. Other uses include pools of internal employees or departments. Each *search pool* record which is found to match the document text is considered as a candidate.

You can store search pools locally, created from either a CSV file or a database source, or use a remote service to create and maintain the *search pools*.

9.17.2 About Remote Matching Service

The Remote Matching Service (RMS) is an option to perform the initial Associative Search Engine match evaluation on a remote service. The remote service imports and configures the pool thereby freeing up local WebCenter Forms Recognition resources and allowing for larger pools than the Brainware V2 and V3 engines. Projects with pools containing more than 500,000 entries may benefit by using this RMS option.

The remote service performs the import, configuration, and update of the pool. Designer displays relevant pool settings, such as Search, Filter, ID and Vendor Type. However, you cannot modify these settings within Designer.

All pools used in RMS are alphanumeric. Therefore you cannot modify the Enable alphanumeric identifier option.

9.17.3 Configure Associative Search Engine

To configure the Associative Search engine, complete the following steps:

1. Switch to *Definition Mode*.
2. In the left pane, on the **Fields** tab, select a field.
3. In the right pane, on the **Analysis** tab, from the **Available Analysis Engines** list, select **Associative Search Engine**.

Note: Until the engine is configured correctly, a red circle and a corresponding error message displays on the General tab.

4. On the **General** tab, from the **Used core engine** list, select the required engine.

- If you selected **Brainware V2** or **Brainware V3**, configure a local search pool.
 - If you selected **Remote Matching Engine**, configure the connection to a Remote Matching Service.
5. On the **General** tab, the **Class name format** is used for **Supervised Learning Workflow (SLW)**. If you do not use SLW, type the class name of any one column from the table above in square brackets.
- [ID]
6. In the **Field contents format** field, type the text you want to appear in the ASE field when extraction is successful.

Note: The text may contain multiple column names, as listed in the table above, but must include at least one. Each column name must be enclosed in square brackets and can be broken across multiple lines. Any text outside of square brackets is considered fixed text and is included whenever extraction is successful.

[ID] [Name] [Street]
[City], [State], [ZIP]
Tel: [Phone]

Result: The format is correct if the entry box's background color turns to white. If the entry box's background color remains red, click the **Check Format** button and examine the error message at the top. If the engine is configured correctly, a green circle and a corresponding message displays on the **General** tab.

7. Optional. Configure the advanced settings.

9.17.3.1 Create a Local Search Pool

You can create a local search pool based on either the contents of a CSV file or the results of query against an existing ODBC connection.

To create a local search pool, complete the following steps:

1. On the **Import** tab, under **Import source**, select one of the following options.
 - Import from a semicolon-separated CSV file
 - Import from ODBC data source
2. Optional. If you use Runtime Server to update the search pools, select **Automatic import in Runtime Server enabled**.
3. If you selected **Import from a semicolon-separated CSV file**, complete the following substeps:
 - a. In the **CSV file** field, type or browse to the desired CSV file.

Note: For more information, refer to [Section 9.17.3.2: Import Data from a CSV File](#).

- b. Optional. If the CSV file is not located in the project directory, clear **Path relative to project file**.
- c. If the CSV file contains a header row, select **Import field names from first row**.

Note: The table on the General tab uses the information from the header row to create the column names. If the CSV files does not contain a header row, the table uses the names from the previous import or generic column names.

4. If you selected **Import from ODBC data source**, complete the following sub steps:
 - a. From the **ODBC** source list, select the desired source.

Note: For more information, refer to [Section 9.17.3.3: Create an ODBC Data Source](#).

- b. In the **User name** field, type the user name.
 - c. In the **Password** field, type the password.
 - d. In the **SQL query** field, type the query that selects the data to be used in the search pool.
5. Under **Import destination**, select **Save supplier pool in separate file** and then, in the field below, specify the directory in which you want to place the search pool files. The import function creates the specified directory if it does not exist.

Note: The Save Supplier Pool inside Project option is not recommended due to performance reasons.

6. Optional. If this directory is not located in the project directory, clear **Path relative to project file**.
7. In the **Pool filename without extension** field, type a filename. This applies to all files for this search pool.

Note: Search pool files for more than one ASE field can be located in the same directory. However, you need to indicate a unique filename for each such field.

8. Click the **Import** button.
9. On the **General** tab, in the **ID** column, select a unique identifier.
10. If the identifier is alphanumeric, select **Enable alphanumeric identifier**.
11. In the **Search** column, select the fields you want to search for.
12. On the **Import** tab, click **Import**.

9.17.3.2 Import Data from a CSV File

The CSV data source must contain at least 3 records or line items, and each record may contain up to a maximum of 3070 characters. Values within each record must be semicolon separated and each record must contain the same number of records. Optionally, the CSV file may contain a header row, which also must contain the same number of values.

The CSV data source must include at least one ID column for which each record has a unique value. The values in this column can be either purely numeric or alphanumeric.

The maximum file size for a CSV file is 150 MB.

Note: CSV import files that contain text in Chinese, Japanese, Korean, Thai, or extended languages such as Russian, must be available in the UTF-16 Little Endian format and you must use the core engine Brainware V3.

9.17.3.3 Create an ODBC Data Source

You can use a query on an external database to create the search pool.

The data returned by the query must include at least 3 records and at least one ID column for which each record has a unique value. The values in this ID column can be either purely numeric or alphanumeric.

The database must be accessible as an ODBC data source. To create an ODBC data source, complete the following steps:

1. Open Windows **Administrative Tools** and then double-click **ODBC Data Sources (32-bit)**.
2. In the **ODBC Data Source Administrator (32-bit)** dialog box, on the **System DSN** tab, add the data source.

9.17.3.4 Configure a Local Search Pool

Once a local search pool has been built, you must configure which data columns are used for searching, filtering and/or identification.

To configure a local search pool, complete the following steps:

1. On the **General** tab, in the **ID** column, select the appropriate ID data column.
2. In the **Search** column, select the data columns which you want to search against.
3. Optional. The **Brainware V3** engine allows for multi-column attribute filtering. To use this script-based filtering of the search pool, select **Filter** for each data column you want to filter on.

Note: Only select data columns which you want to use for filtering. Selecting all columns can have adverse effect on performance.

4. Optional. If applicable, select **Vendor type** for the data column which contains the corresponding values.

Note: Vendor type is used for Supervised Learning Workflow (SLW).

5. On the **Import** tab, click the **Import** button.

Result: If the local search pool has been correctly configured, a green circle displays at the top of the General tab with message indicating that the engine is ready.

6. To test the extraction with the open document corresponding to the data search pool, on the toolbar, click the **Classify / analyze current document** button and observe the value extracted to the field.

9.17.3.5 Configure a Remote Matching Service

To use an existing search pool on a remote matching service, complete the following steps:

Note: The General tab displays the data column configuration as defined within the remote matching service. This configuration is not editable from within Verifier. The alphanumeric identifier option as well as all Import tab options are also disabled. Any changes to Search, Filter, ID, and/or Vendor settings for the data columns must be made within the remote matching service.

1. On the **General** tab, from the **Used core engine** list, select **Remote Matching Service**.
2. Enter the **URL** and **Connection String** provided by the remote matching service.
3. Click **Reconnect**.
4. From the **Search Index** list, select a pool.

9.17.3.6 Configure the Advanced Settings

Advanced settings enable you to control which document text is searched, establish validation thresholds and enable secondary candidate evaluation. To configure the advanced settings, complete the following steps:

1. On the **Advanced** tab, verify or modify the settings for **Threshold and Distance**.
2. Under **Search Areas**, use the controls to define the specific regions of the document from which text is searched. Complete the following substep:
 - Within each tab, **First Page**, **Subsequent Pages**, and **Last Page**, you can define up to two regions by entering **Left**, **Top**, **Width** and **Height** values. These values represent a percentage of the page size. The graphic at the right of each tab illustrates the defined regions.

Note: For single-page documents, only the First Page regions are applied. For a two-page document, First Page and Last Page regions are applied. Only for documents with 3 or more pages the Subsequent Pages regions are applied to all but the first and last pages.

3. By default, the **Enable candidate evaluation using** check box is selected. This activates the secondary candidate evaluation. This feature generally enhances candidate weighting. The document text which is used for secondary candidate evaluation can be set to match that defined within **Search Areas**, the text appearing on the first page of the document, or all document text. Select one of the following options:

- Above configured search areas.
 - First page only.
 - All pages of document.
4. Address Analysis is an alternative to the default secondary candidate evaluation. To configure address analysis, complete the following substeps:

Note: Using this alternative is not recommended.

- a. Clear **Enable candidate evaluation using**.
- b. Select **Enable address analysis**.
- c. If the phrase must exist on the document, select the **Phrase is mandatory** check box.
- d. In the table below, in the **Phrase** column, type the phrase which you expect to find near the candidate text.
- e. In the **Status** column, select **Incl.** to include the candidates which appear near the phrase or select **Excl.** to exclude such candidates.

Note: Enable candidate evaluation using and Enable address analysis are mutually exclusive. If one of them is enabled, this automatically disables the other one.

9.17.4 Multi-Column Attribute Filtering

Both the Brainware V3 core engine and the Remote Matching Service options support multi-column attribute filtering. By default, the Associative Search Engine performs a search against all search pool records. The multi-column Filter attribute can limit the search to records with specific attribute values, such as **Washington** for city. Attribute values are case-sensitive.

You must specify attribute values for each Filter column in script. For example, you can use this feature to filter Verifier Vendor search results or within the field PreExtract event to limit allowable RTS Vendor values. First use the *ClearFilterAttributes* method to clear all attributes of a previous search, and then configure the filters using the *AddFilterAttribute* method. Only one filter attribute may be filtered on multiple values (OR operator). Subsequent filter attributes are connected with the AND operator.

For more information, refer to *SCBCdrSupExSettings* methods and properties in the *WebCenter Forms Recognition Scripting Guide*.

9.17.5 Disable Duplicate Record Suppression for Local Search Pools

By default, a record which is very similar to a previous record in the same search pool is not considered when searching. Use the registry key value *ASEnginePoolAllowedCharDifference* to enable the ASE engine to return candidates that the engine perceived as duplicates of existing records. To create the registry key value, complete the following steps:

Note: This method applies only to local search pools. For a remote matching service, refer to the documentation for the specific matching service.

1. Launch the Windows registry editor.
12. Complete one of the following substeps:
 - For a 32-bit machine, navigate to [HKEY_LOCAL_MACHINE\SOFTWARE\Oracle\Cedar].
 - For a 64-bit machine, navigate to [HKEY_LOCAL_MACHINE\SOFTWARE\Wow6432Node\Oracle\Cedar].
13. In the right pane, right-click and then click **New > DWORD (32-bit) Value**.
14. In the **Name** field, type **ASEnginePoolAllowedCharDifference** and then click **OK**.
15. Right-click the **ASEnginePoolAllowedCharDifference** key and then click **Modify**.
16. In the Edit DWORD (32-bit) Value dialog box, in the **Value data** field, to enable the feature type **0** (zero).
17. Re-analyze the document.
18. Optional. To disable the feature, delete the registry key **ASEnginePoolAllowedCharDifference**.

9.17.6 Use the Associative Search Engine in Verifier

You can use the Associative Search Engine in Verifier to enable the user to select the correct result from candidates. Complete the following steps:

1. Switch to *Verifier Design Mode*.
2. Create a default form for the class, if not yet available.
3. If required, increase the field size.
4. Add a button.
5. Define an action.
6. On the toolbar, click the **Show/hide** script button.
7. In the Script View window, click **Edit** **References**.
8. In the References dialog box, ensure that the **Cedar Associative Search Engine Library** is selected and then click **OK**.
9. Modify the *Document_OnAction* event in the script according to your needs.

9.18 About the Brainware Table Extraction Engine

Use the Brainware Table Extraction (BTE) engine to extract columns and cell data from tables. Further, the engine can interactively train table extraction line-by-line or row-by-row. This engine can learn which lines to extract, which lines to ignore, and which lines to use for learning. The engine learns line types, color codes them, and recognizes when the learnset changes. To use BTE, all documents must have been scanned to a resolution of 300 dpi.

The Generic Table Extraction, included with the BTE, provides additional extraction configuration. For more information, refer to [Section 9.18.6.3: Configure Generic Table Extraction](#).

Use the Brainware Table Extraction Engine for the following purposes:

- Reduce the time required to configure data extraction from tables compared to a scripted solution
- Reduce the time required for table verification in Verifier
- Improve table data recognition rates

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

- Learning lines
- Learning column mapping

9.18.1 Learning Lines

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

Primary Line	A line that defines the table structure. The BTE engine applies advanced and precise similarity analysis for all primary lines. All primary lines must be well structured and similar to each other in many of the rows to be extracted. However, the engine easily supports an unlimited number of different types of primary lines for one table definition. The primary line must be the first line in the table row and must contain at least four words.
Secondary Line	A line between primary lines. The engine applies smooth similarity analysis for these types of lines, which is possible because BTE searches only between two neighboring primary lines. This allows BTE to extract data that varies widely, which often happens with multi-line descriptions. There is no limitation on the number of words in secondary lines, and no limitation on the number of secondary lines. However, a document's page must have at least one primary line; otherwise the engine does not extract secondary lines from this page.
Wrong Line	A primary line that is learned as a negative line sample. In other words, no lines classified by the engine as a member of one particular <i>wrong</i> line class will be extracted. In principle, it is possible to learn an unlimited number of wrong lines, though this will take effect only during in-document learning. Cross-document learning, that means learning the whole document after all fields are completely valid, may not automatically train the wrong lines.

It is possible to learn an unlimited number of different line types. However, the overall quality of the extraction may suffer when learning too many lines.

You can apply the learning of lines in the lines learning or lines highlighting mode.

To learn a line the engine requires the line to contain a minimal number of words. This value defines the number of words that makes a line eligible for learning.

9.18.2 Learning Column Mapping

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.

Likewise, 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 through the Switch Table Highlighting [Ctrl + Q] menu option in Verifier or through pop-up menu options Show Lines and Show Columns of the document viewer in Verifier or Designer.

The user can map the column data in the lines in column mapping learning or column highlighting mode. The user can switch between the learning (highlighting) modes through the Switch Table Highlighting (Ctrl + Q) menu option in Verifier or through pop-up menu options Show Lines and Show Columns of the document viewer in Verifier or Verifier Train Mode in Designer.

9.18.3 Pre-Train Brainware Table Extraction

To build a pre-learnset that improves the performance when working with large document volumes, you can extract a table before the actual training.

Complete the following steps:

1. Switch to *Definition Mode*.
2. From the **Options** menu, select **Pre-train Brainware Table Extraction**.

9.18.4 Configure Brainware Table Extraction

To set up a table field and define column names for *Brainware Table Extraction*, complete the following steps:

1. Switch to *Definition Mode*.
2. On the **Classes** tab on the left side of the screen, select a parent class.
3. On the left side of the screen, click the **Fields** tab.
4. Insert a new field. Right-click anywhere on the tab, select **Add Field Definition** from the popup menu, and then give the new field a name.
5. Right-click the field and select **Table** from the popup menu to change it from a text field to a table field.
6. Right-click in the table field and select **Show Properties** from the popup menu. The properties pane displays on the right.
7. On the **Analysis** tab, under **Available Analysis Engines**, select **Brainware Table Extraction**. The settings for the *Brainware Table Extraction Engine* are displayed.
8. On the **Columns** tab, insert the columns to include in the extracted table. To do this, click the **Insert Columns** button.

9. Enter a name for the column.
10. Check the **Column Required** check box if the column must not contain null values.
11. Select the **Invalid if empty**, if the entries are required for each cell of the column.
12. Check the **Multiline Cells** check box if the engine should not convert multiline text to a single line.
13. Clear the **Visible** check box if you want to hide the column.

Note: The invisible columns must be valid to validate the entire table.

14. Repeat steps 8 - 12 to add more columns as required.
15. To delete a column, click the **Delete columns** button.
16. To move a column up or down, click the arrow buttons.
17. To use the Generic Table Extraction, on the **Labels & Formats** tab, insert a label.

Note: The Generic Table Extraction becomes automatically activate as soon as you configure a column label. For more information, refer to [Section 9.18.6.3: Configure Generic Table Extraction](#).

18. To insert a column format, complete the following sub steps:
 - a. On the **Labels & Formats** tab, click the **Insert formats** button.
 - b. In the **Formats** column, type the expression. For more information, refer to [Section 14.1: What are Regular Expressions?](#)
19. From the **Compare Type** list, select a type.

Note: For more information, refer to [Section 9.20: About the Format Analysis Engine](#).

20. On the **Advanced** tab, in the **Minimal number of words** field, type the number of words according to the demands of the project documents.
21. Under **Primary lines extraction**, in the **Minimal threshold** field, modify the value, if for example the correct primary line was not extracted.
22. Under **Secondary lines extraction**, in the **Minimal threshold** field, modify the value, if for example the correct secondary lines were not extracted.

9.18.5 Use the Derive Options in Child Classes

To use the same column definitions as in the parent class, use the derive options in child classes. Complete the following steps:

1. In the left pane, on the **Classes** tab, select a derived class.
2. In the left pane, on the **Fields** tab, select a table.
3. In the right pane, on the **Analysis** tab, on the Columns tab, complete the following sub steps:

- a. Select **Derive column names** to use the same set of columns as the parent class.

Note: This enables the Derive column definition and Derive column validation formats options and disables all controls related to adding, deleting, renaming and sorting columns.

- b. Optional. Select **Derive column definition** to inherit column settings.
- c. Optional. Select **Derive column validation formats** to inherit validation settings.

9.18.6 Generic Table Extraction

The generic mode of *Brainware Table Extraction* performs without prior scripting or training. The extraction is based on the matching of column labels and additionally works with simple mathematical correlation between columns. Furthermore, it increases the speed of the extraction process.

9.18.6.1 Column Labels

Generic Table Extraction identifies the table header line with the help of the column labels. The column headers that match with defined **Column Labels** will be mapped. Matches will be placed in the data table under the column name that is defined in the column list.

The search for column labels is a *fuzzy* search providing some results even for misspelled or poorly OCR'd words. To ensure that the engine finds adequate matches, define all column headers that you expect in the documents as column labels of the corresponding column name.

To ensure that the engine finds adequate matches, define all column headers that are expected in the documents as **Column Labels** of the corresponding **Column Name**. For example, for the column name **Quantity** define the column labels **Amount**, **Qty**, **Quantity**, **Quantity Shipped**, etc.

9.18.6.2 Numeric Correlation of Columns

The *Numeric Correlation* uses the mathematical interdependency of values to assign the correct values to the corresponding columns. The first three columns with the numbers **0**, **1** and **2** in the column list on the **Column** tab will be checked for correlation.

The correlation follows the scheme: **0** × **1** = **2**.

A common example for an *Invoices* project is *Single Price* × *Amount* = *Total Price*. The columns need to be ordered in the correct sequence, for example:

1. Single Price.
2. Amount.
3. Total Price.

Note: Use this feature only when you expect the columns to correlate in most Documents

9.18.6.3 Configure Generic Table Extraction

To use the *Generic Table Extraction* feature, complete the following steps:

1. In the right pane, on the **Labels & Formats** tab, from the **Column Name** list, select a column.
2. Complete the following sub steps:
 - a. To insert a column label and activate the generic table extraction, click the **Insert labels** button.
 - b. Type a unique name for the label with at least three characters.
 - c. Repeat the previous steps for all required column labels.
3. Optional. To enable correlation, complete the following sub steps:

Note: For more information, refer to [Section 9.18.6.2: Numeric Correlation of Columns](#).

- a. On the **Advanced** tab, select the **Correlation (enable/disable)** check box.
 - b. On the **Columns** tab, move the three columns selected for correlation to the top of the list followed by all other columns.
4. Optional. Delete all **Column Labels** to turn off the Generic Table extraction.

9.18.6.4 Limitations

- The user cannot control the identification of a label line (table header line) or the selection of the columns that are going to be mapped to the data table through a threshold approach. The confidence of their detection cannot be controlled by the threshold settings.
- To learn a line, the engine requires the line to contain a minimal number of words. The **Minimal Number of Words** value defines that minimal number of words that makes a line eligible for learning. Set this value according to the demands of the project documents.
- The label line expressions can consist of multiple words, but there is no difference between specifying two expressions, for example **line** and **number**, and one single expression **line number**.

9.18.7 About Display Names for Table Columns

You can specify user-friendly display names for table columns. You can also use this feature for dynamic translation of verification forms into different languages.

The following areas on the UI use these names:

Note: If you are working with table columns in scripts, you must use the original column names.

- Verification Forms
- Verifier's Verifier Test Mode
- Verifier's Verifier Train Mode
- Verifier

- The column mapping control in context of Brainware Table Extraction interactive learning.
- The column mapping and swapping operations in Verifier.

9.18.7.1 Specify Display Names for Columns

Designer provides settings to set up custom column names for verification table objects. To set the desired table column display names, complete the following steps:

1. Switch to *Definition Mode*.
2. In the left pane, on the **Classes** tab, select a class.
3. Switch to the *Verifier Design Mode* and select the desired table object.
4. Right-click the table and select **Properties** from the popup menu. The *Forms Designer* displays the table object's properties.
5. Click **Column Settings**.
6. To modify the desired column names, select the column row and under the **Display Name** enter the required column name.
7. Click **OK**.

9.19 About the Check Analysis Engine

The Check Analysis engine is a combined analysis and evaluation engine which extracts fields specific to checks without requiring a prior OCR.

The engine cross-validates the Courtesy Amount Recognition and Legal Amount Recognition amounts, detecting possible mismatches. The Check Analysis engine also provides signature-presence verification. For each extracted field the check analysis provides a position and a number of weighted candidate values or texts.

9.19.1 Configure the Check Analysis Engine

To configure the Check Analysis Engine, complete the following steps:

1. Switch to *Definition Mode*.
2. In the left pane, on the **Fields** tab, select a field.
3. In the right pane, on the **Analysis** tab, from the **Available Analysis Engines** list, select **Check Analysis Engine**.
4. From the **Field Type** list, select a field type. For more information, refer to [Section 9.19.2: Field Types](#).
5. On the **Check Analysis** tab, modify the values for **Threshold and Distance**. For more information, refer to [Section 9.1: About Threshold and Distance](#).

9.19.2 Field types

The Check Analysis Engine extracts the following field types from a check Document.

9.19.2.1 MICR (Magnetic Ink Character Recognition) Line

Extracts the bank, account and check number information line from the bottom of the check.

9.19.2.2 Courtesy Amount Recognition (CAR)

Returns the numeric check amount as a dollar amount.

9.19.2.3 Legal Amount Recognition (LAR)

Returns the text check amount as a dollar amount.

9.19.2.4 Cross-Checked Amount

The check amount as a dollar amount, determined using the CAR and LAR results.

9.19.2.5 Check Date

Returns the check date.

To improve extraction accuracy, the engine captures documents with check dates up to 12 months in the past and one month in the future based on the current system date.

The engine extracts the check date in the format *MMM dd yyyy*, such as Jan 07 2016. If necessary, you can process out of date checks using the *PCDateHint* property in a script.

For more information, refer to *PCDateHint* in the *WebCenter Forms Recognition Scripting Guide*.

9.19.2.6 Check Number

The check number, returning the position which is generally in the upper right.

9.19.2.7 Remitter Location

If the engine detects a candidate, the field value is [Location found]. For more information, refer to [Section 9.19.5: Extract the Remitter Information](#).

If the engine cannot detect the remitter position on the document, a Verifier user can add this information manually during verification.

9.19.2.8 Payee

Returns the payee information as text.

To boost the recognition accuracy of payee lines on checks, the engine works with a predefined vocabulary list which comprises the following parameters. You can combine either PayeeName or Keyword with Weight. It is not possible to use both in the same project.

9.19.2.8.1 Payee field settings

Payee name: Defines the already known potential payee candidates. Use the GUI table or script to provide the candidates list.

Keyword: Define keywords in order to get a subset of candidates. Keywords can be single words or parts of the payee name. Use the GUI table or script to provide the candidates list. For details how to use a keyword dictionary, refer to [Section 9.19.4: About Vocabulary Keyword Search](#).

Weight: This value reflects the occurrence likelihood for a vocabulary entry. The value of the weight parameter ranges from 0 (zero) to 15. Typically, most words have a weight of zero. A weight of 15 corresponds to a vocabulary entry that appears in approximately 50% of the images. Use the GUI table or script to define the weight values.

Vocabulary Coverage: This value reflects the expected percentage of checks that match the predefined names. To set the value, use the slider. For details how to set the coverage in script, refer to *PayeeVocCoverage* in the *WebCenter Forms Recognition Scripting Guide*.

The combination of the parameters has considerable effect on Payee recognition accuracy.

For more information on configuration of vocabulary parameters, refer to *PayeeVocEntries* and *PayeeVocCoverage* in the *WebCenter Forms Recognition Scripting Guide*.

You can also define threshold and distance values for the Payee field. For more information, refer to [Section 9.1: About Threshold and Distance](#).

9.19.2.9 Signature and Signature 2

The engine returns the position of one or two signatures on the check.

If the engine detects at least one signature, the Signature field value is set to Signature 1 and returns the position of the first signature. The position of the second signature, if present, is returned as a second candidate.

The second candidate can also be accessed using the Signature 2 field. If this field is included and the engine detects two signatures on the check, the Signature 2 field value is set to Signature 2 and returns the position of the second signature.

All signature candidates have equal weighting.

9.19.3 Payee Vocabulary List

You can define the candidate lists using the table on the Check Analysis tab or by script.

9.19.3.1 About Vocabulary List Duplicates

If you add duplicates to the Vocabulary table, refreshing deletes them. The first entry persists.

If the table contains duplicates created by the user and the script, both entries display in the table. The extraction uses the entry created by the script, and its weight value overrides the entry created by the user.

If the script adds duplicates, the validation uses the final one added.

9.19.4 About Vocabulary Keyword Search

The Keyword Only check box allows you to restrict the search to keywords. This enhances the search performance. In order to use this option, define a keywords dictionary in the script.

For details how to create keyword entries, refer to *PayeeVocEntries* in the *WebCenter Forms Recognition Scripting Guide*.

9.19.4.1 Define the Vocabulary List in the UI

To define the vocabulary list in the UI, complete the following steps:

1. On the **Analysis** tab, click **Add**.
2. In the **Payee Name** column, type the payee name.

Note: If you don't change the default value **Vocabulary entry [X]**, the next refresh action deletes the entry.

3. To move to the **Weight** field, press [**Ctrl + Right Arrow**].
4. In the **Weight** column, type a value.

Note: The maximum value is 15.

5. Optional. To insert a new row at the bottom of the table, click **Add**.
6. Optional. To insert a new row above the selected row, click **Ins**.
7. Optional. To delete the selected row, click **Del**.

9.19.4.2 Define the Vocabulary List using a Script

To add script entries to the vocabulary list, complete the following steps:

1. Enable the reference to the **Cedar ParaCheck Analysis Engine Library**. For more information, refer to *Enable Payee Line script methods* in the *WebCenter Forms Recognition Scripting Guide*.
2. In the script, define the entries using the method `AddItemToVocabulary`. For more information, refer to *AddItemToVocabulary* in the *WebCenter Forms Recognition Scripting Guide*.
3. Execute the script in your project.

Note: Candidates added by a script are read-only and display in red.

4. On the **File** menu, click **Save Project** in order to keep the project entries generated by the script.

9.19.5 Extract the Remitter Information

To extract the remitter information using secondary OCR, complete the following steps:

1. Switch to *Definition Mode*.
2. In the left pane, on the **Fields** tab, select the remitter field.
3. In the right pane, on the **Fields** tab, select **Read field again**.

9.20 About the Format Analysis Engine

Use this engine to extract data located at arbitrary positions in the documents.

Format analysis uses a formal description of the character string you are looking for. For each field, a set of search patterns can be defined. All strings within the document text that match at least one of the specified patterns are candidates. Typically, there are several candidates.

Format analysis is usually combined with an evaluation engine. That is, the user can take a couple of examples and select the correct candidate from the possible ones. The engine thereby learns to select automatically. Alternatively, you can combine format analysis with custom evaluation methods implemented using scripts.

It is the primary analysis method and can be used with unstructured documents and with structured documents such as forms.

Format analysis is a search within the document text for strings that match certain predefined patterns. Configuring this method requires that you define the patterns, such as, the format strings.

Note: Performance can be improved by applying templates or identical settings to multiple fields.

9.20.1 Compare methods

The string you are looking for may consist of only one word only or several words. Whether the string is finally found in the document also depends on rules as to how it is constructed from words. Default rules are set for this and should work in most cases. If your format string looks fine, but you still cannot find candidates, you may have to adjust them. For more information, refer to [Section 9.20.7: Define the Rules for String Construction from Words](#).

The engine provides the following methods to find strings:

9.20.2 String Compare

String Compare is a very simple search method that finds each literal occurrence of the specified format string.

9.20.2.1 Levenshtein

Levenshtein search is an error-tolerant search method that finds each literal occurrence of the specified format string, but also strings that can be derived from the specified one by inserting, interchanging or deleting single characters. The number of key operations required to derive the erroneous string determines whether there is still a match.

Use this method to account for typical typographical errors like character interchange.

Example

Zero Errors	1 Error	2 Errors
Invoice	invoike	invoke
	involve	involve

9.20.2.2 Trigram

Trigram search is another error-tolerant search method. To compare two words, they are fragmented into groups of three characters called trigrams. The number of identical groups determines whether there is still a match.

Use this method to account for OCR errors in your document.

Example

Word	1 st Trigram	2 nd Trigram	3 rd Trigram
brain	bra	rai	ain
train	tra	rai	ain

9.20.2.3 Simple Expression

This is the default method that you can use to specify simple format patterns. In simple expressions, some characters have a special meaning. All other characters have no special meaning and represent themselves.

Character	Description
#	Represents any number. For example: ### matches 123
@	Represents any upper-case or lower-case letter. For example: @@@@ matches Love @# matches U2
?	Represents any alphanumeric character. For example: ?rain matches brain and train
' '	Indicates the start or end of a word. For example: 'ABC' matches ABC , but not ABCD
[]	Indicates a number or range of repetitions of the previous character. For example: #[3] matches any three-digit number like 123 a[1,2] matches a and aa me[1,2] matches met and meet

9.20.2.4 Regular Expressions

Use regular expressions to precisely specify complex format patterns. Verifier supports a subset of regular expressions. For more information, refer to [Section 14.1: What are Regular Expressions](#).

9.20.3 Designator Types

If you specify a Designator type, the *Format Analysis Engine* looks for your search string, but the candidate is a different string that has a geometrical relationship to the search string.

For example, you can search for the string **Phone** to extract the subsequent information, such as a phone number.

The following Designator types are available:

Word Below Returns the word below the search string.

Word Above Returns the word above the search string.

This Block Returns the block that contains the search string.

Previous Word Returns the word left to the search string.

Next Word Returns the word right to the search string.

End of Line Returns all words between the search string and the end of the line.

Next Line Returns the line below the search string.

Next Paragraph Returns the paragraph below the search string.

Next Block Returns the block below the search string.

9.20.4 Define Format Strings

You can define a set of format strings for each field. Before you begin, consider the following information:

- Review your documents and, for each field, create a list of the strings that you want the engine to extract. This way makes it easier to identify the patterns to search for.
- Do not try to find precisely one candidate, try to generate a set. Learning works better with both correct and false examples.
- Try to anticipate other likely formats. If the correct candidate has no chance to be identified, the need for manual correction increases.

9.20.5 Configure the Format Analysis Engine

To configure the Format Analysis Engine, complete the following steps:

Note: Applying a template or identical settings to multiple fields shortens processing time.

1. Switch to *Definition Mode*.
2. Optional. To make it easier to visualize this task, load a document.
3. On the **Analysis** tab, select the **Format Analysis** tab at the bottom.
Initially, most of the controls in this tab are disabled. You need to specify the format string before you can adjust the search method.
4. In the **Format Strings** table, the last row is always empty. Click the row number and type a search string. Press the [Enter] key to add a new row. The string remains selected. You can now make further adjustments.
5. If required, select a different search method from the **Compare Method** list box. By default, **Simple Expression** is set.
6. If required, select the **Find Designator** and from the **Designator Type** list, select a type. For more information, refer to [Section 9.20.3: Designator Types](#).
7. Under **Prefix** and **Suffix**, you can enter characters or syllables that are part of a word, but are ignored when found at the beginning or the end of the word.
For example, you can look for **22,99**, and **\$22,99** will match.
8. Under **Ignore Characters**, you can enter characters that are part of a word, but are ignored when found anywhere within the word.
For example, you can look for **213-4789**, and **2134789** will match.
9. To define additional strings for the current field, repeat the steps above. Note that you can combine search methods freely. However, a word can only be member of one candidate. The format strings will be evaluated in the same order as listed in the *Format Strings* table. To change the order, click a row number and drag the corresponding string to its new position. To delete a string, click the row number and press the [Delete] key.
10. Initially, most of the controls in this tab are disabled. You need to specify the format string before you can adjust the search method.

11. In the **Format Strings** table, the last row is always empty. To enter a new search string, click in this row number and type a search string.

9.20.6 About Testing

- Test the analysis step with candidate highlighting to check early whether you have defined the suitable format strings. However, be careful if you use field inheritance. The extraction test always involves a classification step. You can easily end up changing the child class settings by mistake. Always check the class name that displays in the bottom-left edge of the window. This is the class you are currently editing.
- *Format Analysis* normally yields several candidates. Therefore, a candidate evaluation must take place.

9.20.7 Define the Rules for String Construction from Words

The *Format Analysis Engine* may incorrectly fragment an OCR'd candidate into several words, or concatenate several words. Therefore, the string search also covers combinations and subsets of words. The corresponding rules are valid for all format strings assigned to a given field.

9.20.7.1 Define Rules for Words

To define rules for string construction from words, complete the following steps:

1. Switch to *Definition Mode*.
2. In the left pane, on the **Fields** tab, select a field that uses the **Format Analysis Engine**.
3. In the right pane, on the **Analysis** tab, on the **General** tab, set the rules as needed. For information about each rule, [Section 9.20.7: Define the Rules for String Construction from Words](#).

9.20.7.2 Rules for String Construction from Words

This topic explains the rules for string construction from words.

Max. Compare Distance

This feature allows you to extract a search string from a word, neglecting a certain number of characters at the beginning and at the end. The compare distance is computed as follows:

Compare Distance = $1.00 - \text{Number of Characters (Search String)} / \text{Number of Characters (Word)}$

A match requires that the actual compare distance is less than or equal to the maximum compare distance. For example:

Search String: **1234** (4 characters)

Max. Compare Distance: **0.30**

Word 1: **\$1234** (5 characters)

Compare Distance 1: $1.00 - (4/5) =$
0.20

This is a match as $0.20 < 0.30$

Word 2: **\$1234.00** (8 characters)

Compare Distance 2: $1.00 - (4/8) =$
0.50

This is not a match as $0.20 >$

0.30

Max. Word Count

Allows you to extract strings that consist of multiple words.

By default, the hyphen character is a word separator. Therefore, the phone number **123-456-7890** consists of three *words*. To get a match in the *Format Analysis*, a maximum word count of **3** or more is required.

Max. Gap Between Words

Specifies the maximum distance in millimeters that permits word concatenation during the search. Note that the requirements strongly depend on font size.

Max. Candidate Length

Specifies the maximum length in millimeters that a candidate is allowed to be. A candidate that exceeds the maximum length is not accepted.

Compare Case Sensitive

Select this option to make the candidate search case-sensitive.

Keep Spaces Between Connected Words

By default, the engine deletes spaces between connected words. To keep the spaces, select this option and type the spaces in your format string.

9.20.8 Restrict the Analysis to Specific Areas

You can restrict the text that is analyzed to specific areas of the document. This may increase processing speed and accuracy.

To restrict the analysis to certain areas within the documents, complete the following steps:

1. Optional. To make it easier to visualize this task, load a document.
2. Switch to *Definition Mode*.
3. In the left pane, on the **Fields** tab, select a field that has format analysis defined.
4. On the **Analysis** tab, select the **Regions** tab at the bottom.
5. Select the **Restrict Engine to Region** check box.
6. Select the pages that are affected by the restriction.
7. For each selected page, enter the regions that are to be taken into account.

9.21 Table Analysis Engine

Use this rule-based engine to extract data from tables.

The recommended engine to extract table data is the *Brainware Table Extraction* engine.

As with Brainware Table Extraction, traditional table analysis yielded several candidates, but the one with the highest confidence is used automatically.

Therefore, no evaluation step is required. You can use scripts to access the remaining candidates.

In principal, there are two approaches to traditional table analysis.

Fixed Layout	If the position of table elements is constant, you can use their coordinates to locate the table and its elements. This resembles the <i>Zone Analysis</i> of text fields.
Variable Layout	If the position of table elements varies, or to analyze different table layouts with a single set of settings, you can use column labels and formats to locate a set of table candidates. This resembles the <i>Format Analysis</i> of text fields.

You can also mix both approaches.

The *Table Analysis Engine* allows you to extract a single table from each document. This table can span multiple pages, provided that the table layout does not change over the entire range. To be accessible by table analysis, tables must consist of the following elements.

- Columns
- Rows
- Cells

Limitations

- Reading tables with interleaved columns is not supported.

- Reading tables with empty rows is only supported if the table has a footer line.

9.21.1 Switch between the Table Analysis Engine and the Brainware Table Extraction Engine

You can convert from the Table Analysis Engine to the Brainware Table Extraction, as all column names and definitions remain the same. To convert a table from the Table Analysis Engine to the Brainware Table Extraction, complete the following steps:

1. Switch to *Definition Mode*.
2. In the left pane, on the **Fields** tab, select the **Table** field you want to convert.
3. In the right pane, on the **Analysis** tab, select **Brainware Table Extraction**.
4. If necessary, remove redundant extraction-related script code for the converted table field.

Note: If a table cell has multiple extracted lines, use the *RowNumber* property to determine the row number for each table line extracted with the Brainware Table Extraction engine in all validation and extraction scripts. For more information, refer to *RowNumber* in the *WebCenter Forms Recognition Scripting Guide*.

9.21.2 About Settings Table

To configure the table analysis, you need to define a settings table. This is a table containing the columns you want to retrieve from the document, along with a number of properties for each column. The tables within the document are called the *document tables*.

Table analysis performs the following actions:

- Identifies the document table that best matches the settings table.
- Analyzes the rows and columns of the document table.
- Maps the columns of the document table to those of the settings table.
- Extracts the content of the document table.

Each part of the analysis can create its own set of candidates. In the user interface, only the table candidate with maximum confidence is returned so that learning is not required. However, the workdoc contains a list of all table candidates, sorted by confidence. These can be accessed from scripts.

The first step in defining new table settings is to specify the columns to be searched for. Each column has to be assigned a name, which can be used to access the column data from script (alternatively to use the column index), and a set of properties. Defining tables for a derived class enables the check boxes. Column names and column labels can be different from each other.

9.21.3 Define the Settings Table

To define the table columns, complete the following steps:

1. Optional. To make it easier to visualize this task, load a document from the respective class.
2. Switch to *Definition Mode*.
3. In the left pane, on the **Fields** tab, select a field.

4. In the right pane, on the **Analysis** tab, from the **Available Analysis Engines** list, select **Table Analysis Engine**.
5. On the **Column** tab, complete the following sub steps:
 - a. Click the **Insert columns** button.
 - b. In the **Name** column, type a name for the column.
 - c. Optional. If the column must not contain null values, select the **Column Required** check box.
 - d. Optional. If entries are required for each cell of the column, select the **Entry Required** check box.
 - e. Optional. If the engine should not convert multiline text to a single line, select the **Multiline Cells** check box.
 - f. Optional. If you do not want to display the column in Verifier, clear the **Column Visible** check box. You can set this property in a script to display the table column only when it contains a special content. For more information, refer to *ColumnVisible* in the *WebCenter Forms Recognition Scripting Guide*.
 - g. Optional. To delete a column, click the **Delete columns** button.
 - h. Repeat the previous steps for all required columns.
 - i. Optional. To move a column up or down, click the arrow buttons.

Note: During verification of tables, the user navigates from invalid cell to invalid cell by pressing Enter. If no entries are required in a column, empty cells are valid, even though the document table contains corresponding input. To avoid losing information while keeping the ability to accept empty cells, you can make input mandatory for a column, but permit forced validation in the verification step. This is a pre-validation option. If it is selected for a table cell, it returns an invalid table cell when no entry is found. The result of the pre-validation will be discarded when the table cell is set to valid within the script.

9.21.4 Column Labels and Formats

For each column, you must to specify a label that the engine uses to identify the column.

Optionally you can specify format strings to determine whether the column cells have a valid content.

In most cases, all column labels are positioned at about the same horizontal position, means they belong to one text line. This line is called *label line*. In some cases, the labels extend over more than one text line. The engine can process single-line as well as multi-line labels.

The label line plays a central role for the table analysis, to detect the following information:

- The start position of the table
- The preliminary table layout and
- The preliminary mapping of the table columns.

To properly assign a cell to a column, some degree of overlap with the label must exist. If a label overlaps with cells of different columns, the column is created using the cells with the most overlap.

9.21.5 Define Column Labels and Formats

For multi-line labels, no special settings are required as they are automatically combined to one label during table analysis.



To define column labels and formats, complete the following steps:

1. In the right pane, on the **Labels & Formats** tab, from the **Column Name** list, select a column.
2. Click the **Insert label** button.
3. In the **Label** column, type a name for the label.
4. Repeat the previous steps for all required labels.
5. Optional. To delete a column, click the **Delete columns** button.
6. Optional. To specify column formats, complete the following sub steps:
 - a. Click the **Insert format** button.
 - b. In the **Formats** column, type the format string.
 - c. In the **Compare Type** column, select a type. For more information, refer to [Section 7.17.5: Compare Methods](#) and [Section 14: Regular Expressions](#).
 - d. Repeat the previous steps for all required format strings

9.21.6 Define Header and Footer Lines

To define table headers and footers, complete the following steps:

1. On the **Analysis** tab, click the **Header & Footer** tab.
2. In the **Insert / Delete Labels** list, define a set of labels using the following buttons:

Button	Description
	Adds a new entry at the bottom of the list. At the insertion point, type a name for the entry.
	Deletes the selected entry without displaying a confirmation message.

3. In the **Insert / Delete Footers** list, define a set of labels using the same buttons.

9.21.7 About Column Layouts

For each table, specify the algorithms used to determine the column layout. The following options are available:

- Layout analysis, based on the labels of the label line and the entries of the individual rows.
- Fixed layout, using positions for the left and right border of each column.

9.21.8 Define the Column Layout

To define the column layout, complete the following steps:

1. On the **Analysis** tab, select the **Column Layout** tab.
2. Complete one of the following options:
 - For variable layout, select the **Variable Column Position** option. Optionally, you can check **Create New Table Candidates**. This is useful when a column is not represented by a label. However, it implies the creation of a large number of candidates, thus effecting the time required for the table analysis.
 - For fixed layout, select the **Fixed Column Position** option. In the table at the bottom of the tab, specify the distance of the left and right column borders from the left border of the page in millimeters. These ranges are highlighted in the document area.

9.21.9 Determine the Top of the Table

For each table, you need to specify the method you want to use to search for the top. Complete the following steps:

1. In the right pane, on the Top of Table tab, complete any of the following steps:
 - Select **Search label line**.
 - Select **Search header**.

Note: Use this setting only in combination with a fixed column layout.

- Select **Use fix position** and then, in the **First page** and the **Subsequent pages** fields, specify the distance from the top of page in millimeters. If you analyze the document, the specified position highlights in the document area. Use this setting only in combination with a fixed column layout.
 - Select **Search edges** to enable the Table Analysis Engine to find table columns by searching for edges. As it finds columns, you would select the columns that you defined for the table on the Column tab. To make this mapping, use either script, or for invoices, the Column Mapping tab that has mapping with fields that are predefined. Within the script, you can navigate through the cells of the columns to check the content. The combo box on this tab contains the column names that are defined on the column tab. You can select the column name that corresponds to the Description, Price, and so on. You can define the table columns that you want to write to the workdoc on the Column tab. With multiple options, the analysis stops when the first criterion is met.
2. Optional. In the **Skip** field, type a number to skip a number of text lines below the tabletop.

Note: Use this feature only for tables with a fixed layout.

9.21.10 About the Bottom of a Table

For each table, you need to specify the algorithms used to search the bottom.

The following options are available:

- Searching for variations in the text line distance.
- Searching a graphical line.
- Searching the footer based on the list of footer strings defined. For more information, refer to [Section 9.21.6: Define Header and Footer Lines](#).
- Searching for empty cells in a control column.

You can also combine these methods.

9.21.11 Determine the Bottom of a Table

To determine the bottom of a table, complete the following steps:

- On the Analysis tab, on the End of Table tab, select any of the following options:

Note: With multiple options, the analysis stops when the first criterion is met.

- Select **Change of Textline Distance** if the table ends when the text line distance changes by more than 50%.
- Select **Textline Distance Larger** and type a value into the associated text box if the table ends when the text line distance exceeds a predefined threshold.
- Check **Horizontal Lines** if a horizontal line indicates the end of the table.
- Check **Detected Table Footer** if one of the footers defined previously indicates the end of the table. This option must be checked if your tables contain empty rows.
- Check **Missing Entry in Control Column** if an empty cell in specified columns indicates the end of the table. In the table at the bottom of the tab, mark the Used check box for at least one column.

9.21.12 About Graphical Lines

In some cases, tables contain vertical or horizontal graphical lines. Horizontal lines can be used to detect the end of a table instead of or in addition to further criteria. Vertical lines can indicate the segmentation of columns, but are currently not evaluated.

9.21.13 Comment Lines

Comments, such as free text in tables, are automatically ignored if two requirements are met:

- There are at least two lines with cells above the comment.
- There is at least one line with cells below the comment.

Otherwise, comment lines will interfere with the table analysis.

9.21.14 Detect Comment Lines

To detect comment lines, complete the following step:

- On the **Analysis** tab, on the **Comment lines** tab, select any of the following options:

Note: With multiple options, the analysis stops when the first criterion is met.

- Select the **overlapping column borders** if entries spanning multiple columns indicate a comment. In the table at the bottom of the tab, mark the **Used** check box for at least one column.
- Select the **empty cells in control columns** if empty cells in specified columns indicate a comment. In the table at the bottom of the tab, mark the **Used** check box for at least one column.

9.21.15 About Field Inheritance

To allow a coupling between parent and child settings, the concept of *inheritance* has been introduced. Simply stated, child classes inherit all fields from their parents. If the *Table Analysis* engine is selected for the table field in the child class, a copy of the parent settings is created, which may be further modified. For this purpose, the table settings have been subdivided into three groups.

Column Settings The definition of the number of columns and their names.

Extraction Settings All parameters controlling the extraction algorithms.

Validation Settings The definitions of formats for the individual columns.

Depending on the inheritance settings, the corresponding controls in the table settings tabs will be disabled.

Note: If a child class is created manually, without passing the *Supervised Learning Workflow* process, it is necessary to activate the *Brainware Table Extraction* engine for the *Table* field at the child class level.

9.21.16 Using Derive Options in Child Classes

To use the same column definitions as in the parent class, use the derive options in child classes. Complete the following steps:

1. On the **Analysis** tab, select the **Columns** tab.
2. Check **Derive Column Names** to use the same set of columns as the parent class. All controls related to adding, deleting, renaming and sorting columns will now be disabled. The **Derive Column Definition** and **Derive Column Validation Formats** check boxes will be enabled.

Note: This enables the Derive column definition and Derive column validation formats options and disables all controls related to adding, deleting, renaming and sorting columns.

3. Optionally, select **Derive Column Definition** to inherit column settings.
4. Optionally, select **Derive Column Validation Formats** to inherit validation settings.

9.21.17 Column Mapping

After you select the columns from which you would extract information for the workdoc in the **Columns** tab, you can search for certain table columns in a batch. This is most commonly used if you are working with invoices. To search for columns, click **Search Edges** in the top of **Table** tab and then use the *Column Mapping* feature in *Table Analysis*. For more information, refer to [Section 9.21.9: Determine the Top of the Table](#).

9.21.18 Map Table Columns

To map table columns, complete the following step:

1. On the **Analysis** tab, on the **Column Mapping** tab, select the required column from the **Description**, **Quantity**, **Single Price**, and **Total Price** lists.

9.21.19 Map Table Columns in a Script

Another possibility for defining table columns is to define the mapping within a script.

For more information, refer to *SCBCdrTable* in the *WebCenter Forms Recognition Scripting Guide*.

9.22 Zone Analysis Engine

Zone Analysis Engine is suitable for extracting data located at fixed positions within the documents, such as with classical forms.

9.22.1 About the Zone Analysis Engine

Use this method to extract data located at fixed positions in a document. This requirement is usually met with classical forms. For each field, define a reading zone or a set of reading zones on the document. The recognition methods available for zone reading include OCR engines, barcode recognition, and optical mark detection (OMR). Typically, zone analysis yields a single candidate, but with multiple reading zones per field, a set of candidates can be obtained. Zone analysis only requires the evaluation step if there are multiple candidates.

You can then use either Brainware evaluation or custom evaluation methods in scripts.

9.22.2 About Reading Zones

Reading zones are fixed areas in documents that contain information you want to extract. They display as transparent rectangles with a gray background and red borders. Each zone receives a name composed from the word *Zone* and a consecutive number.

The following types of reading zones are available:

OCR	The contents of the reading zone are determined using optical character recognition. The textual content of the zone is the result of this process.
Barcode	The contents of the reading zone are determined using barcode recognition. The number or strings represented by the barcode is the result of this process.

OMR	The contents of the reading zone are determined using optical mark recognition. The degree of blackness within the zone is used as indicator. A Boolean value is the result of this process.
-----	--

9.22.2.1 Create Reading Zone

To create a reading zone, complete the following steps:

1. Load a document that belongs to the respective class.
2. Switch to *Definition Mode*.
3. In the left pane, on the **Fields** tab, select a field.
4. In the right pane, on the **Analysis** tab, from the **Available Analysis Engines** list, select **Zone Analysis Engine**.
5. In the middle pane, complete one of the following sub steps:
 - To create an OCR reading zone, click the **Insert OCR zone** button.
 - To create an OMR reading zone, click the **Insert OMR zone** button.
 - To create a barcode reading zone, click the **Insert barcode zone** button.
6. On the document, drag to mark the required zone.

Note: The zone must be large enough to hold the expected content, but not much larger.

9.22.2.2 Adjust Reading Zone

To adjust a reading zone, complete the following steps:

1. Optional. To make it easier to visualize this task, load a document.
2. Complete any of the following steps.
 - Drag the zone to another place.
 - Resize the zone by dragging the corner handles.
 - To adjust the zone size to the size of the entire page, right-click the zone and then click **Fit to image size**.
 - To delete a zone, select the zone and then press **Del**.

9.22.3 About Zone Properties

The zone properties affect extraction steps using zone analysis. Locally, they override project-level settings. You can apply different settings to each reading zone.

Modify the zone properties if you are experiencing recognition-related problems to extract the correct data.

At the zone level, you can use all preprocessing and recognition settings that improve your results. Due to the small reading area, the loss in performance is insignificant.

9.22.3.1 Modify Zone Properties

To modify the zone-level settings, complete the following steps:

1. Load a document from the respective class that has zones.
2. Switch to *Definition Mode*.
3. In the left pane, on the **Fields** tab, select a field that uses the Zone Analysis Engine.
4. In the middle pane, select a reading zone.
5. Right-click the zone and then click **Properties**.
6. In the Zone Properties dialog box, modify the settings. For more information, see the following topics.
 - Modify the settings on the General tab
 - Modify the settings on the Preprocessing tab
 - Modify the settings on the Recognition tab
 - Assign an anchor to a reading zone.

9.22.3.2 Modify Settings on General Tab

To edit size and position of the reading zone, complete the following step:

Note: If multiple reading zones are mapped to a single data field, a candidate evaluation must take place. This usually involves learning. If you change the properties of your reading zones after the learning, you lose all trained knowledge for the extraction.

1. In the **Zone Properties** dialog box, on the **General** tab, complete any of the following sub steps:
 - In the **Name** field, type a new name for the zone.
 - From the **Units** list, select the unit you want to use for geometric parameters, such as pixel, inch, or mm.
 - To adjust the height of the zone to the height of the page, select Fit to image height.
 - To adjust the width of the zone to the width of the page, select **Fit to image width**.
 - To set the distance of the reading zone to the left edge of the page, modify the value in the **Left** field.
 - To set the distance of the reading zone to the top edge of the page, modify the value in the **Top** field.
 - To set the width of the reading zone, modify the value in the **Width** field.
 - To set the height of the reading zone, modify the value in the **Height** field.
 - To enable automatic fitting, select **Enable automatic fitting** and then modify the values in the **Left**, **Right**, **Top**, and **Bottom** fields.
2. If learning was processed before, repeat the training.

9.22.3.3 Modify Settings on the Preprocessing Tab

To modify the preprocessing settings, complete the following steps:

1. Right-click the zone and then click **Properties**.

2. In the **Zone Properties** dialog box, on the **Preprocessing** tab, under **Available Methods**, select any of the following methods and then click the arrow button to move it to the **Selected Methods** list.
 - **Binarisation**: Select this method to convert images to black-and-white images.
 - **Box and Comb Removal**: Select this method to remove black boxes or regular and irregular combs.
 - **Clean Border**: Select this method to improve the quality of images with a dark border as, for example, often generated in photo copies.
 - **Despeckle**: Select this method to remove speckles.
 - **Invert**: Select this method to revert white-on-black pages to black-on-white pages.
 - **Lines Manager**: Select this method to remove horizontal or vertical lines from your document.
3. Optional. To disable a preprocessing method, under **Selected Methods**, select an entry and then click the arrow button to move it to the **Available Methods** list.
4. Optional. To order the preprocessing methods, under **Selected Methods**, select an entry and then click the arrow up or down button.

Note: The methods perform in the listed order.

5. If you selected **Box and Comb Removal**, complete the following sub steps:
 - a. In the **Selected Methods** list, select **Box and Comb Removal**.
 - b. On the **Box or Comb Removal** tab, modify the settings.
6. If you selected **Lines Manager**, complete the following sub steps:
 - a. In the **Selected Methods** list, select **Lines Manager**.
 - b. Optional. On the **General** tab, to remove horizontal lines, select **Remove hor. lines** and then type a minimum line length in millimeters in the **line - horizontal** field.
 - c. Optional. To remove vertical lines, select **Remove ver. lines**. Type a minimum line length in millimeters in the **line - vertical** field.
 - d. Optional. Select **Repair Characters**.
7. To test the effect of the methods, complete the following sub steps:
 - a. Ensure that the current document is visible and not hidden behind the dialog box.
 - b. To test the selected methods, click the **All Steps** button.

Note: The required time displays below the button. In case of performance issues, check the sum of recognition time and preprocessing time. Even if preprocessing takes a while, it may accelerate recognition.

- c. View the document to check the preprocessing effects.

- d. Optional. If you want to repeat the test, click the **Reset** button before you proceed.

9.22.3.4 Modify Settings on the Recognition Tab

To modify the recognition settings, complete the following steps:

1. On the **Recognition** tab, under **Recognition** Type, select **OCR**, **OMR**, or **Barcode**.
2. From the **Available Engines** list, select the recognition engine.
3. To use the selected engine as default engine for the current recognition type, select **Use as default engine**.
4. In the lower pane, modify the engine-specific settings.

9.22.3.5 Zone Testing

To test the zone, complete the following steps:

Note: Avoid tests with entire pages.

1. In the **Zone Properties** dialog box, on the **Test** tab, under **Zone Names**, select a zone if none is selected.
2. Click **Read Zone**.
3. In the **Reco Result** field, observe the result.

9.22.4 About Anchors

It is quite common that paper documents are scanned at an angle. From page to page, there will be fluctuations of the page position. However, zone analysis can only be successful if the relative position of reading zones on documents is constant over the entire document set. So-called anchors can be used to look for prominent geometric features on the document, such as the page corners. After they are found, anchors can provide a fixed coordinate system for the reading zones, i.e. reading zone positions are determined relatively to anchor positions. This can balance fluctuations of several millimeters.

Anchors are special types of reading zones. They look for prominent geometric features on documents, such as page corners. After these features are found, anchors can provide a fixed coordinate system for the reading zones. Position for reading zones is determined relative to anchor positions. This can overcome fluctuations in the page position of scanned images.

Anchors display as transparent rectangles with a gray background and blue borders. Each anchor receives a name composed from the word Anchor and a consecutive number.

9.22.4.1 Create an Anchor

To create and use anchors, complete the following steps:

1. Load a document that belongs to the respective class.
2. Switch to *Definition Mode*.
3. In the middle pane, to create an anchor, click the **Insert anchor** button.
4. On the document, drag to mark the anchor zone.

Note: Ensure that the anchor zone is proportional to the geometric feature that you want to locate. For example, if all the documents in your class contain a 50 mm line in the footer area, your anchor should not have a side length of 5 or 250 mm. The anchor position should be close to the geometric feature you want to use.

5. Optional. To create more anchors, repeat the previous steps.

9.22.4.2 Modify Anchor Properties

To modify the anchor properties, complete the following steps:

1. On the document, right-click the anchor and then click **Properties**.
2. In the **Anchor Properties** dialog box, complete any of the following sub steps:
 - In the **Name** field, type a new name for the anchor.
 - From the **Units** list, select the unit you want to use for geometric parameters, such as pixel, inch, or mm.
 - To set the distance of the anchor to the left edge of the page, modify the value in the **Left** field.
 - To set the distance of the anchor to the top edge of the page, modify the value in the **Top** field.
 - To set the width of the anchor, modify the value in the **Width** field.
 - To set the height of the anchor, modify the value in the **Height** field.
 - To specify for which shape the anchor is looking for, from the **Anchor Type** list, select a type and then modify the properties of the anchor type.
 - Click the **Learn** button and then click **OK**.

9.22.4.3 Assign an Anchor to a Reading Zone

To assign an anchor to a reading zone, complete the following steps:

1. On the document, right-click a zone and then click **Properties**.
2. In the **Zone Properties** dialog box, on the **Anchor** tab, move one or more anchors from the **Available Anchors** list to the **Used Anchors** list.

Note: Only used anchors provide a geometrical reference for the current reading zone.

3. Optional. Repeat the previous steps for other zones.

9.22.5 Assign Reading Zones to a Field

To actually extract data from reading zones, you need to assign them to fields, complete the following steps:

1. Load a document that belongs to the respective class.
2. Switch to *Definition Mode*.
3. In the left pane, on the **Fields** tab, select a field.

4. In the right pane, on the **Analysis** tab, from the **Available Analysis Engines** list, select **Zone Analysis Engine**.
5. On the **Zone Analysis** tab, complete the following sub steps:
 - a. In the **Zone Name** column, select a zone from the list. In the **Page Number** column, specify whether the reading zone is on the first or the last page.
 - b. To create multiple candidates from the zone analysis, repeat the previous step using the next table rows.

Note: If you map multiple reading zones to a single data field, a candidate evaluation must take place.

10 Validation

10.1 About Validation

WebCenter Forms Recognition's robust validation and output formatting feature enables you to standardize the development and maintenance cycles for each implementation of WebCenter Forms Recognition. However, standard validation has not replaced scripted validation. Validation scripts are still supported. For a brief introduction to script validation, refer to [Section 10.9.14: About Validation Scripts](#). For detailed information about scripting, refer to the *Oracle WebCenter Forms Recognition Scripting User's Guide*.

Validation is a quality assurance task that involves confirming whether a field extraction is correct or incorrect. Validation is configured in Designer's *Validation Editor*, but the actual work is done in Runtime Server, Verifier, and Advanced Verifier.

Runtime Server loads the project at runtime and executes validation after extraction but before exporting the project. Documents that fail validation in Runtime Server are forwarded to Verifier for quality assurance. Verifier loads the settings and uses property sheets to display the validation settings.

Validation uses WebCenter Forms Recognition's hierarchical organization of relationships between parent and child elements.

While Runtime Server is processing a batch or list of batches, each containing several documents, it first tries to classify the processed document. After classification, the extraction is performed for each field defined for this document class. The validation rules for a field can either be defined for this document class but can also be inherited from the parent class.

A document will be valid when all fields on the document are valid. If one field is invalid according to standard validation or script validation rules, the whole document is invalid. The script validation for a field is performed only when the standard validation is valid. If one document of a batch is invalid, the whole batch is passed to the Verifier for manual intervention.

10.2 About Standard Validation Engine

The Standard Validation Engine enables you to define validation rules for the project, documents, classes, and fields. The engine must be enabled in order for any validation to occur, or for you to establish validation settings.

10.3 About Validation Settings

Validation properties are inherited along the parent-child hierarchy. Children automatically inherit their parent's validation settings. You can accept inherited settings or override them.

10.4 About Force Validation Feature

Force validation allows you to force a field to be valid, even if it is invalid by definition. If force validation is permitted, the entity being validated is always valid,

regardless of its content or specific settings. Default force validation means that the validation settings are taken from the parent.

10.4.1 Force Validation Settings

Default	Inherits the force validation setting from the parent class or the project.
Permit	Permits to force the validation for the document, class, or field.
Forbidden	Forbids to force the validation for the document, class, or field.

10.4.2 Modify Force Validation Setting for a Project

The force validation setting of the project defines the default force validation setting for classes and fields.

To modify the setting, complete the following steps:

1. Switch to *Definition Mode*.
2. In the left pane, on the **Classes** tab, select the project.
3. In the right pane, on the **Advanced** tab, under **Force Validation**, select either **Permitted** or **Forbidden**.

10.5 Derived Validation

Derived validation tells a child field, document, or class to inherit all validation settings from its parent. You can use derived validation only on trained classes.

Validation rules apply cumulatively from parent to child and not only for the affected class node.

To apply document level validation rules for inherited classes, turn on the **Use derived validation** option for the required derived classes. By default, this option is active for all newly created subclasses.

10.6 About Validation Rules

Validation rules enable you to compare one field to another, establish data criteria, and perform certain mathematical and logical functions.

You can create validation rules for classes as well as for the following field types:

- Text
- Amount
- Date
- Check box
- List
- Table

10.7 About Document Validation

You can validate single documents using a script.

For more information, see *Validate* in the *Oracle WebCenter Forms Recognition Scripting User's Guide*.

10.8 About Validation Rules for Classes

To modify the validation settings for a class, complete the following steps:

1. Optional. To make it easier to visualize this task, load a document.
2. Switch to *Definition Mode*.
3. In the left pane, on the **Classes** tab, select a class.
4. In the right pane, on the **Validation** tab, under **Available Validation Engines**, select **Standard Validation**.
5. Under **Force Validation**, select **Default**, **Permitted**, or **Forbidden**. For more information, refer to [Section 10.4.1: Force Validation Settings](#).
6. Optional. From the **Available Templates** list, select a template and then click **Copy Template**.
7. Optional. For subclasses, select **Use derived validation**. For more information, refer to [Section 10.5: About Derived Validation](#).
8. Optional. On the **General** tab, to build an expression, complete the following sub steps:
 - a. To copy the field name to the expression, double-click an amount field.
 - b. Click one of the function buttons. For more information, refer to [Section 10.8.3: Functions for Validation Rules](#).
 - c. Click one of the operator buttons. For more information, refer to [Section 10.8.2: Operators for Validation Rules](#).
 - d. If required, type the parameters.
 - e. To add the expression to the list of validation rules, click **Add**.
 - f. To add another expression, click **Clear** and then repeat the previous steps.
9. Optional. To delete a validation rule, in the **Validation Rules** list, select the rule and then click **Delete**.
10. Optional. To delete all validation rules, click **Delete All**.

10.8.1 Modify Validation Rule for a Class

To modify a validation rule, complete the following steps:

1. Switch to *Definition Mode*.
2. In the left pane, on the **Classes** tab, select a class.
3. On the **General** tab, complete the following sub steps:
 - a. From the **Validation Rules** list, select a validation rule.
 - b. Modify the rule.
 - c. Click the **Update** button.

10.8.2 Operators for Validation Rules

Operator	Description
=	Equal to.
>	Greater than.
<	Less than.
*	Multiplied by.
/	Divided by.
+	Plus.
-	Minus
(Open group for expression.
)	Close group for expression.
!	Not
!=	Not equal to.
&&	AND.
	OR.

10.8.3 Functions for Validation Rules

Function	Description
SUM	Sums all table rules for a column.
MAX	Sets the maximum value for a column.
MIN	Sets the minimum value for a column.
COUNT	Counts the number of rows in a column.
AVG	Averages table rows.
ABS	ABS function.

10.8.4 About the ABS Function

The ABS function evaluates the absolute value of the enclosed expression and simplifies the expressions.

You can simplify the expression `((NetAmount+VAT-Total)<0.02)&&((NetAmount+VAT-Total)>(-0.02))` by using the ABS function to `ABS(NetAmount+VAT-Total)<0.02`

When using constant values in an expression, such as `ABS(NetAmount+VAT-Total)<0.02` which always use a period as decimal separator and do not use thousand separators.

10.9 About Possible Limitations for Validating Fields

The list below provides an overview of possible limitations when validating fields:

- Handling of OCR rejects.
- Permitted number of characters.
- Range of valid characters.
- Restriction to uppercase / lowercase characters.
- Handling of forced validation.
- Custom validation rules implemented as scripts.

10.9.1 About Delayed Validation

The **Allow delayed validation** feature enables you to perform the validation of fields together at the end of the validation process instead of validating each field separately. This feature increases the validation performance in Verifier and Web Verifier by reducing the number of server requests.

10.9.2 Modify Validation Rules for a Field

To modify the validation settings for a field, complete the following steps:

1. Switch to *Definition Mode*.
2. In the left pane, on the **Fields** tab, select a field.
3. In the right pane, on the **Validation** tab, under **Force Validation**, select **Default**, **Permitted**, or **Forbidden**. For more information, refer to [Section 10.4.1: Force Validation Settings](#).
4. Optional. From the **Available Templates** list, select a template and then click **Copy**.
5. Optional. If the field belongs to a subclass, select **Use Derived Validation**.

Note: This makes all other settings unmodifiable.

6. Optional. To make the field always valid regardless of its content or validation settings, select **Always Valid**.
7. Optional. Select **Allow delayed validation**.
8. From the **Validation Type** list, select a type.
9. Complete one of the following sub steps:

- For a text field, modify the general validation rules for a text field and modify the character filtering settings.
- For an amount field, modify the general validation rules for an amount field, modify the character filtering settings, and modify the output formatting settings for a date or an amount field.
- For a date field, modify the general validation rules for a date field, modify the character filtering settings, and modify the output formatting settings for a date or an amount field.
- For a check box field, modify the general validation rules for a check box field and modify the character filtering settings.
- For a list field, modify the general validation rules for a list field and modify the character filtering settings.

10.9.3 Modify General Validation Rules for a Text Field

To modify the general validation rules for a text field, select **General** tab on the right pane, complete any of the following steps:

- To convert multi-line data to a single, non-delimited line, select **Multi-Line**.
- To allow fields without content, select **Allow empty Field**.
- To define a minimum length for the field, select **Min. Len** and then type the length in the field below.
- To define a maximum length for the field, select **Max. Len** and then type the length in the field below.

Note: For more information about the settings, refer to [Section 10.9.11: General Validation Settings for a Text Field](#).

10.9.4 Modify General Validation Rules for an Amount Field

To modify the general validation rules for an amount field, complete the following steps:

1. Optional. In the right pane, on the **General** tab, select **Amount Range** and then specify the minimum and the maximum amount in the corresponding fields.

Note: For more information about the settings, refer to [Section 10.9.11: General Validation Settings for an Amount Field](#).

2. From the **Region** list, select a region.
3. Optional. Complete the following sub steps:
 - a. Select the required settings from the lists in the right column.
 - b. To copy all settings selected in the right column to the left column, click the << **All** << button.
 - c. Optional. In the **Name** field left to the << **All** << button, type a name for the rule.
4. Optional. Complete the following sub steps:

- a. Select an entry from one of the lists in the right column and then click <<.
- b. Optional. Select another entry from the same list and then click <<.

Note: Use this feature to add multiple allowed values for one setting, for example if you want to allow both “,” and “.” as a **1000 separator**.

- c. Repeat the previous steps for all settings you want to add to the rule.

Note: If you don't select a value for “Currency Symbol” and “1000 separator”, the value “no” is inserted for these fields and amounts without a currency symbol or a decimal separator are validated.

- d. Optional. In the **Name** field left to the << **All** << button, type a name for the rule.
5. Optional. Type additional values in the fields in the left column using the “|” sign to separate the values.
6. To create the rule, click the **Add** button.

Note: If you did not apply any specific values in the left columns, all options available in the lists in the right column are accepted as valid.

7. Optional. To add another formatting, click **New / Clear** and then repeat the previous steps.
8. Optional. To modify an existing formatting, select an entry from the list, modify the settings and then click **Update**.
9. Optional. To deactivate a rule, clear the corresponding check box in the table.

10.9.5 Modify General Validation Rules for a Date Field

To modify the general validation rules for a date field, complete the following steps:

1. Optional. In the right pane, on the **General** tab, select **Date Range** and then specify the date range in the corresponding fields.
2. From the **Region** list, select a region.
3. Complete one of the following sub steps:
 - To add all formats for the region, click **Add All**.
 - To add a specific format, select an entry from the Format list and then click **Add**.
4. Optional. Select **No Future Date Permitted**.

10.9.6 Modify General Validation Rules for a Check Box Field

To modify the general validation rules for a check box field, complete the following steps:

1. Optional. In the right pane, on the **General** tab, in the **Check box Caption** field, type a free text check box caption, such as Yes/No, Accept/Decline, or True/False.

2. In the **Checked Value** field, type the value for a selected check box according to the value you specified in the **Check box Caption** field.
3. In the **Unchecked Value** field, type the value for a cleared check box according to the value you specified in the **Check box Caption** field.
4. Complete one of the following sub steps:
 - To specify whether a check box is selected by default, select **Checked**.
 - To specify whether a check box is cleared by default, select **Unchecked**.

10.9.7 Modify General Validation Rules for a List Field

To modify the general validation rules for a list field, complete the following steps:

1. Optional. In the right pane, on the **General** tab, to define a minimum length for the field, select **Min. Len** and then type the length in the field below.
2. Optional. To define a maximum length for the field, select **Max. Len** and then type the length in the field below.
3. To add a value to the list, click the **Insert** button and then type the value in the list field below.
4. Repeat the previous step for all required entries.
5. Optional. To allow only values from the list, select **Allow list values only**.
6. Optional. To populate the list with the results from the Associative Search Engine (ASE), select **Fill list of Suppliers**.

Note: The Remote Matching Service does not support this option.

7. To automatically select the most matching entry from the list while typing in Verifier or Verifier Test Mode, select **Show Nearest Value**.
8. Optional. To sort the items on the list in alphanumeric order, select **Sort list items**.

10.9.8 Modify General Validation Rules for Table Fields

To modify the validation settings for table fields, complete the following steps:

1. In the left pane, on the **Fields** tab, select a table field.
2. In the right pane, on the **General** tab, complete the following sub steps:
 - a. To specify the type the engine uses to validate the table field, in the **Type** column, select an entry from the list.
 - b. Optional. To assign a template to the field, in the **Template** column, select an entry from the list.

10.9.9 Predefined Replacement Characters

Use this feature to replace characters that the engine often misinterprets during OCR. The following replacement characters are predefined for date and amount fields.

You can add new characters to the list, or modify the existing ones.

Character to replace	Replacement Character
O	0
B	8
i (lower case I)	1
l (lower case L)	1
I (upper case I)	1

10.9.10 Modify Character Filtering Settings

To modify the character filtering settings, complete the following steps:

1. Under **Valid Characters**, complete one of the following sub steps:

Note: NOTE: These settings are not available for amount fields.

- To allow all numbers, click the **All Numbers** button.
 - To allow all characters, click the **All Characters** button.
 - To allow all numbers and all characters, click the **All** button.
 - To allow only specific characters, click one or more of the buttons representing the character you want to allow.
2. Optional. In the **Remove Characters** field, type the characters you want to remove.

Note: Amount and date fields provide the following predefined list of characters to be removed. You can modify this list.?: ;" ' & ? ~ ! @ # % ^ *.

3. Optional. Under Replace Characters, complete the following sub steps:
 - a. In the **Old Char** column, type the character you want to replace.
 - b. In the **New Char** column, type the character you want to insert.
4. Optional. Complete the following sub steps:
 - a. To accept all OCR results, regardless of quality, select **No Rejects**.
 - b. To determine the minimum OCR confidence for the content to be validated, in the **OCR Confidence** field, specify a percentage value.

Note: A low confidence level mandates excellent OCR quality.

5. Under Input Conversion, select one of the following options:
 - To prevent conversion, select **No Conversion**.
 - To convert all character in the field to upper case, select **Upper Case**.
 - To convert all character in the field to lower case, select **Lower Case**.

10.9.11 General Validation Settings for a Text Field

Multi-Line Converts multi-line data to a single, non-delimited, line.

Allow Empty Field Defines whether a text field can be empty.

Min. Len Defines the minimum length of the field.

Max. Len Defines the maximum length of the field.

10.9.11.1 General Validation Settings for an Amount Field

Amount Range Establishes the setting for validating a range of values.

Minimum Value Fields with values below this range will not be validated.

Maximum Value Fields with values above this range will not be validated.

Region WebCenter Forms Recognition supports 27 regional settings, derived from default written languages.

Currency Symbol Selections vary based on the **Region** setting. For example, if you chose **U.S. English** for the region, the symbols for U.S. currency are available. If you chose **Canadian English**, the symbols for Canadian currency are used. All currency formats for the region are accepted as valid.

Positive Format Uses the currency settings for the **Region** you selected and standard currency formats for depicting positive numbers.

Negative Format Uses the currency settings for the **Region** you selected and standard currency formats for depicting negative numbers.

Decimal Symbol Decimal separator. This options works only in combination with **Decimal Places**.

Decimal Places Number of digits that may appear after the decimal symbol. This is useful for truncating decimal characters. This options works only in combination with **Decimal Symbol**.

Thousands Separator Thousands separator. This options works only in combination with **Digit Grouping**.

Digit Grouping You can select to have two or three characters per group or not to group digits at all.

10.9.11.2 General Validation Settings for a Check box Field

Check box Caption Enables you to indicate whether a check box should be captioned **Yes / No, Accept / Decline, True / False** or some other value.

Checked Value	Enables you to indicate whether a check box means Yes, True, Accept or some other value when it is checked. This setting should not conflict with Check box Caption .
Unchecked Value	Enables you to indicate whether a check box means No, False, Decline or some other value when it is unchecked. This setting should not conflict with Check box Caption .
Default (Checked / Unchecked)	Enables you to indicate whether a check box should be selected or cleared by default.

10.9.11.3 General Validation Settings for a List Field

Min Len	Establishes a minimum valid field length.
Max Len	Establishes a maximum valid field length.
Insert / Delete List Items	Create or remove list rows.
Allow List Values Only	Only values that appear on the list will be interpreted as valid.
Fill List of Suppliers	This option is only available for fields which are designed to use the <i>Associative Search Engine</i> .
Show Nearest Value	Automatically selects the most matching entry from the list while typing in Verifier or Verifier Test Mode.
Sort List Items	Arranges items in the list in alphanumeric order.

10.9.11.4 General Validation Settings for Table Fields

Type	Indicates whether the table column should be validated as text, amount, check box, list, or date.
Template	Assigns a template to the field.

10.9.12 About Output Formatting

Output formatting enables you to standardize the format for date and amount fields when you export them. You must specifically enable the output formatting at the field level.

10.9.12.1 Modify Output Formatting Settings for a Date or an Amount Field

To modify the output formatting, complete the following steps:

- For more information, refer to [Section 10.9.12.3: Output Format Validation Settings for a Date Field](#) and [Section 10.9.12.2: Output Format Validation Settings for an Amount Field](#).
1. In the right pane, on the **Output Format** tab, modify the settings.

2. To activate output formatting, select **Enable Output Format**.

10.9.12.2 Output Format Validation Settings for an Amount Field

Region	WebCenter Forms Recognition supports 27 regional settings, derived from default written languages.
Currency Symbol	Selections vary based on the Region setting. For example, if you chose U.S. English for the region, the symbols for U.S. currency are available. If you chose Canadian English , the symbols for Canadian currency are used. All currency formats for the region are accepted as valid.
Positive Format	Uses the currency settings for the Region you selected and standard currency formats for depicting positive numbers.
Negative Format	Uses the currency settings for the Region you selected and standard currency formats for depicting negative numbers.
Decimal Symbol	Decimal separator. You cannot use the same symbol as for the digit grouping symbol.
No. of digits after decimal	Number of digits that may appear after the decimal symbol. This is useful for truncating decimal characters.
Digit grouping symbol	Thousands separator. You cannot use the same symbol as for the decimal symbol.
Digit grouping	You can elect to have two or three characters per group or not to group digits at all.

10.9.12.3 Output Format Validation Settings for a Date Field

Region	Provides predefined regional date settings.
Format	Provides predefined regional date formats depending on the selected region.
Sample	Displays how the date appears depending on the regional and format settings.

10.9.13 About Validation Templates

Validation templates enable you to easily apply frequently used validation settings to other fields or classes within the same or other projects.

You can create templates for specific kinds of validation for fields and classes and then apply them to other fields and classes in the project.

Note: Template names are case-sensitive.

10.9.13.1 Create Validation Template for Fields

To create a validation template for fields, complete the following steps:

1. Optional. To make it easier to visualize this task, load a document.
2. Switch to *Definition Mode*.
3. In the left pane, on the **Fields** tab, select a field.
4. In the right pane, on the **Validation** tab, complete the following sub steps:
 - a. From the **Validation Type** list, select a type.
 - b. Optional. From the **Available Templates** list, select a template and then click Copy.
 - c. Modify the settings.
5. Click **Save**.
6. In the **New Template** dialog box, type a name for the template and then click **OK**.

10.9.13.2 Apply Validation Template to a Field

You can apply a field validation template created in another class in your project to any field in any class within the same project. To assign an existing template to a field, complete the following steps:

1. Optional. To make it easier to visualize this task, load a document.
2. Switch to *Definition Mode*.
3. In the left pane, on the **Fields** tab, select a field.
4. In the right pane, on the **Validation** tab, from the **Validation Type** list, select a type.
5. From the **Available Templates** list, select a template and then click **Copy**.
6. If required, modify the settings.

Note: The modifications apply only to the selected field and not to the template.

7. On the **File** menu, click **Save**.

10.9.13.3 Create Validation Template for Classes

To create a validation template for classes, complete the following steps:

1. Switch to *Definition Mode*.
2. In the left pane, on the **Classes** tab, select a class.
3. In the right pane, on the **Validation** tab, complete the following sub steps:
 - a. Optional. From the **Available Templates** list, select a template and then click **Copy Template**.
 - b. Modify the settings.
 - c. Click **Save as Template**.
 - d. In the **New Template** dialog box, type a name for the template and then click **OK**.

10.9.13.4 Apply Validation Template to a Class

To apply an existing template to a class, complete the following steps:

1. Switch to *Definition Mode*.
2. In the left pane, on the **Classes** tab, select a class.
3. In the right pane, on the **Validation** tab, from the **Available Templates** list, select a template and then click **Copy**.
4. If required, modify the settings.

Note: The modifications apply only to the selected class and not to the template.

5. On the **File** menu, click **Save**.

10.9.13.5 Modify Field template

Any changes you make apply only to the template within the project and not to any exported template. To modify a field template, complete the following steps:

1. In the left pane, on the **Classes** tab, right-click the project and then click **Show Validation Templates**.
2. In the right pane, on the **Validation Templates** tab, from the **Field Type** list, select a type.
3. Under **Available Validation Templates**, select a template.
4. Modify the settings on the General and the Char Filtering tab, and if available, on the **Output Format** tab.
5. On the **File** menu, click **Save Project**.

10.9.13.6 Import Field Template

To import a field template, complete the following steps:

1. In the left pane, on the **Classes** tab, right-click the project and then click **Show Validation Templates**.
2. In the right pane, on the **Validation Templates** tab, click **Import**.
3. In the **Open** dialog box, select an EXP file and then click **Open**.
4. In the **Import Templates** dialog box, clear the check boxes for the templates you do not want to import and then click **OK**.

10.9.13.7 Export Field Template

To export one or more templates, complete the following steps:

1. In the left pane, on the **Classes** tab, right-click the project and then click **Show Validation Templates**.
2. In the right pane, on the **Validation Templates** tab, click **Export**.
3. In the **Export Templates** dialog box, clear the check boxes for the templates you do not want to export and then click **OK**.

Note: All selected templates are exported to one EXP file.

4. In the **Save As** dialog box, in the **File name** field, type a name and then click **Save**.

10.9.13.8 Delete Field Template

To delete a template, complete the following steps:

Note: You cannot delete a template that is in use.

1. In the left pane, on the **Classes** tab, right-click the project and then click **Show Validation Templates**.
2. In the right pane, on the **Validation Templates** tab, from the **Field Type** list, select a type.
3. Under **Available Validation Templates**, select a template.
4. Click **Delete**.
5. In the **Delete Template** dialog box, click **Yes**.

10.9.13.9 Rename Field Template

To rename a template, complete the following steps:

Note: You cannot rename a template that is in use.

1. In the left pane, on the **Classes** tab, right-click the project and then click **Show Validation Templates**.
2. In the right pane, on the **Validation Templates** tab, from the **Field Type** list, select a type.
3. Under **Available Validation Templates**, select a template.
4. Click **Rename**.
5. In the **Rename Template** dialog box, type a new name for the template and then click **OK**.

10.9.14 About Validation Scripts

Oracle WebCenter Forms Recognition offers various script methods and properties to further customize validation. Note that script validation overrides the validation settings made in the Oracle WebCenter Forms Recognition Designer. You can apply validation scripts for projects, documents, fields, table cells, table rows, and entire tables.

The automated field formatting occurs after a *[Field]_Validate* event is fired if a field is completely valid. You can access the formatting results in a *FormatForExport* event script. The *FormattedText* property contains the formatted text, while the *Text* property contains the unformatted text.

For more information, refer to *Oracle WebCenter Forms Recognition Scripting User's Guide*.

10.9.14.1 About Output Formatting in Scripts

You can access the formatted value in the *SCBCdrField* property *FormattedText*. The same property is also available for table cells and can be accessed through a property, as shown below:

```
pTable.CellFormattedText(Column, Row)
```

You can access the properties can from script events, such as the *ExportDocument* or *Document_Validate* event.

For more information, refer to *FormatForExport*, *FormattedText*, *ExportDocument*, and *Validate* in the *Oracle WebCenter Forms Recognition Scripting User's Guide*.

10.9.14.2 Value Property for Simplified Scripting of Extended Validation Routines

A *Value* property is available for both table cells and header fields and can be used to significantly simplify access to the automatically extracted data from within custom script. This property contains a field's value in the variable type depending on the field validation type.

The *Value* property is *variant* and returns a date for *date* fields, currency for *amount* fields, a Boolean value for *check box* fields, the index of the list item for the *list* fields or text for the other field types.

The *Value* property for the header fields is accessible as *pField.Value* and for the table cells as *pTable.CellValue (Column, Row)*.

10.9.14.3 Open a Script

To open a script, complete the following steps:

1. Switch to *Definition Mode*.
2. On the toolbar, click the **Show/hide script** button.
3. Modify the script according to your needs.

10.9.15 About the Number of Invalid Fields

This feature allows to control of the number of fields that are set to invalid in case a document level validation expression fails.

By default this option is deactivated, that means if a validation rule like `NetAmount+VAT=GrossAmount` fails, the validation subsystem makes all three fields invalid.

Activate the setting if you want that only one of these three fields is made invalid.

If the option is activated, and a rule with both table columns and header fields, such as `SUM(Table.TotalPrice)=GrossAmount` fails, the validation subsystem attempts to invalidate one of the available header fields in the expression and not the table cells.

When making a table column invalid according to a document level validation rule, the validation engine marks only the first cell of the column as invalid.

10.9.15.1 Control the Number of Invalidated Fields

Use this feature to mark only one field as invalid when validation fails for a document.

To activate the feature, complete the following steps:

1. Switch to *Definition Mode*.
2. In the left pane, on the **Classes** tab, select the project.

3. In the right pane, on the **Advanced** tab, select **When validation rule fails, mark only one field as invalid (otherwise all)**.

11 Evaluation

11.1 About Field Evaluation

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

Oracle WebCenter Forms Recognition supports the following methods for candidate evaluation:

- Evaluation using scripts.
- Evaluation using the Brainware Extraction engine or the Brainware Field Extraction engine.

Setting up the field evaluation is not required, if you use scripts or if you use zone analysis with a 1:1 zone/field relationship.

In all other cases, you must specify an evaluation method. During evaluation, Oracle WebCenter Forms Recognition analyzes the areas around each candidate for characteristic properties that can be learned.

11.2 Configure Brainware Extraction Engine

To configure the Brainware Extraction engine, complete the following steps:

1. Switch to *Definition Mode*.
2. In the left pane, on the **Fields** tab, select a field.
3. In the right pane, on the **Evaluation** tab, under **Available Evaluation Engines**, select **Brainware Extraction**.
4. Optional. On the **Brainware Extraction** tab, adjust the settings for **Threshold** and **Distance**.
5. To specify the geometry of the area around the candidate to analyze, on the **Microlayout** tab, select one of the following options:
 - **Table-based**. The context of the candidate is taken from a top-asymmetric region around the candidate, the line to the left and the column above. This configuration is the best for candidates in tables. This is the default microlayout.
 - **Left / top oriented**. The context of the candidate is taken from a left-asymmetric region around the candidate and the line to the left. This configuration is the best for candidates in captions or after colons.
 - **Symmetric**. The context of the candidate is taken from a symmetric region around the candidate. This configuration is the best for candidates in running text.
 - **Line oriented**. The context of the candidate is taken from the region to the left and to the right of the candidate. This configuration is the best for candidates in lines.

Note: Microlayout settings have only a small influence on extraction results. They need many samples to be distinguished. In most cases, the table-based microlayout setting is the best choice.

11.3 Brainware Field Extraction Engine for Generic Fields Extraction

The *Brainware Field Extraction* (BFE) engine introduces a dimension in automatic indexing of header field data on a generic level, i.e. when multiple logical document classes are handled with a single internal knowledgebase. The engine uses an advanced N-gram *Triton* technology and provides high quality field extraction based on the following statistical evaluation of:

- Surrounding keywords around the learned field.
- Format of the learned field.
- Typical location of the learned field.
- Correlation between different fields learned in context of one document class.

All the information above is learned automatically. The engine also autonomously extracts candidates while time-consuming specification of regular expressions through the *Format Analysis Engine* is no longer required for the processing.

The engine fully supports Supervised Learning Workflow.

The *Brainware Field Extraction* engine can be configured and trained in the same way as the *Brainware Extraction engine*.

The engine can be assigned to any header field of a document class using the **Evaluation** tab of the field definition settings and selecting **Brainware Field Extraction** from the list of available engines.

As for the *Brainware Extraction engine*, settings view also displays the learn status for the selected field on the **General** tab.

11.3.1 Differences between the Brainware Field Extraction Engine and the Brainware Extraction Engine

There are some differences related to configuring of the *Brainware Field Extraction* engine as opposed to the *Brainware Extraction Engine*:

- *Format Analysis* engine pre-processing and time-consuming configuration is not required for the *Brainware Field Extraction* engine function. See the following limitations and refer to [Section 11.3.3: About Brainware Field Extraction candidates in a script](#).
- This *Brainware Field Extraction* engine does not support *microlayout* settings.
- Unless handled otherwise by a script, the *Brainware Field Extraction* engine deletes all previously available candidate objects, such as the ones created by the preceding execution of the *Format Analysis* engine and creates and evaluates self-extracted candidates.

11.3.2 Configure Brainware Field Extraction Engine

To configure the Brainware Extraction engine, complete the following steps:

1. Switch to *Definition Mode*.
2. In the left pane, on the **Fields** tab, select a field.

3. In the right pane, on the **Evaluation** tab, under **Available Evaluation Engines**, select **Brainware Field Extraction**.
4. Optional. On the **General** tab, adjust the settings for **Threshold** and **Distance**.
5. If you want the engine to treat the field as an amount, select **Amount Field**.

Note: The engine considers all amount fields as the ones with potential numeric correlation.

11.3.3 About Brainware Field Extraction Candidates in a Script

The following script sample shows how to replace the extracted *Brainware Fields Extraction* candidates with the ones extracted by the *Format Analysis* engine or how to keep both kinds of candidates in the candidates array.

```
Private Sub Date_PostEvaluate(pField As SCBCdrPROJLib.SCBCdrField,
pWorkdoc As SCBCdrPROJLib.SCBCdrWorkdoc)
    Dim i As Long
    For i = pWorkdoc.Fields.ItemByName("Date").CandidateCount - 1 To 0
Step -1
        ' Comment the code line below if you'd like to keep both types of
candidates (from FAE and BFE)
        pWorkdoc.Fields.ItemByName("Date").RemoveCandidate(i)
    Next i
    pWorkdoc.DocState = CDRDocStateClassified
    pWorkdoc.Fields.ItemByName("Date").FieldState = CDRFieldStateReset
    Project.AllClasses.ItemByName("Invoices").AnalyzeField pWorkdoc,
    "Date"
End Sub
```

11.3.4 Usage and Notes

Use the engine for generic training only, that means for data extraction from unclassified documents. Train the engine with at least 10 documents per extracted field, while the ideal amount of documents per field is 50-100 samples.

We do not recommend to use the *Brainware Field Extraction* engine for the **vendor** level training, as the Brainware Extraction engine achieves perfect extraction results with just a few, that means 1-5, documents for a particular high-volume vendor class.

Even though the engine is integrated in Supervised Learning Workflow, due to the above mentioned reasons, it is advisable to use the *Brainware Extraction Engine* for automatic vendor-level training. For this purpose, it is recommended to assign the *Brainware Extraction Engine* to the header fields of the root classes, using a special dedicated class for generic extraction through the *Brainware Field Extraction* engine. For example, the following classes' hierarchy:

- Invoices
 - Generic
 - VendorClass1
 - ...
 - VendorClassN
- Void

The **Invoices** class is the root template class for Supervised Learning Workflow training, while the **Generic** class defines generic extraction through the *Brainware*

Field Extraction engine. The standard classification result for invoices must be set to **Generic**, so that in case a document has not been classified to one of available vendor classes with precise layout extraction, it goes to **Generic** class with a generic extraction pattern defined. After the generic processing has been applied, the classification result must be set back to **Invoices** in case further Supervised Learning Workflow processing is to be maintained. This can be achieved by placing the following script code into the *PostExtract* event's handler of the **Generic** document class:

```
Private Sub Document_PostExtract(pWorkdoc As SCBCdrPROJLib.SCBCdrWorkdoc)

    pWorkdoc.DocClassName = "Invoices"

End Sub
```

In some project configurations, when vendor classification succeeds but does not deliver any results for some selected fields, it may also be appropriate to take an advantage of the **Retain Previous Extraction Results** option and define double extraction processing to combine generic extraction with vendor level extraction.

11.4 About Training Header Fields in Normal Train Mode

Use this feature for quick learning of Brainware Field Extraction engine without necessity to define header field formats through the Format Analysis engine's regular expressions and without a need to configure Supervised Learning Workflow to use *Verifier Train Mode* for the purpose of Brainware Field Extraction learning.

This feature can be used for training of the former Brainware Extraction engine, as well. On the other hand, unless specifically desired, it is not recommended to utilize it for the Brainware Extraction engine for the following reasons:

- Further extraction with the Brainware Extraction engine requires availability of the expected extraction result among the pre-extracted (through Format Analysis engine) candidates. The engine only evaluates the existing candidates assigning them individual weights. These weights are then used by the general extraction subsystem to assign the final extraction result. Therefore, if extraction result was not among the auto-extracted candidates initially and was just created at runtime during learning, the probability that the same case will occur for a hypothetical similar document to extract is very high. This will then lead to complete unavailability of the correct extraction result.
- When a document is being trained with Brainware Extraction, the engine processes the learned field's content using it as the single member of an internal *Correct Results* class and using all its other available candidates as multiple members of an internal *Wrong Results* class. The further extraction process is close to smart classification between these two classes. Thus, it is clear that overall quality of Brainware Extraction engine's outcome also depends on quality of the *Wrong Results* class content. This content, in case no candidates were available prior to start of the document learning procedure, will be reduced to just one single member for the *Wrong Results* class, which is going to be the first available word in the document. While such content is sufficient for training and extraction in general, it is not optimal in terms of quality.

11.5 Train Header Fields in Normal Train Mode

To train the header fields in the Normal Train Mode without preceding *Format Analysis* processing, complete the following steps:

1. Open an OCRed document that is assigned to the required class.
2. Switch to *Normal Train Mode*.
3. On the toolbar, select a highlighting mode, such as **Highlight Words**.
4. In the left pane, on the **Fields** tab, select a header field that has the Brainware Extraction or Brainware Field Extraction engine assigned.
5. In the middle pane, on the document, drag to select the required words.
6. NOTE: The system adapts the area to the best matching position.
7. To create the candidate and apply the required field assignment for further training, double-click the selected area.

11.6 About Learning the Extraction

You need a couple of sample documents to learn the extraction. In these documents, manually assign some candidates to the respective fields. This enables Oracle WebCenter Forms Recognition to learn how to extract the data.

11.7 About Creating Learnsets

The set of sample documents that you need to provide for data extraction is called the extraction learnset. Carefully select the samples as the quality of your learnset is crucial for the success of the extraction.

When you select the sample documents for the learnsets, consider the following:

- Use only documents with good OCR results for the learnset. When in doubt, highlight the OCR results to review the documents.
- Never use handwritten documents.
- Use only documents that generate candidates for all relevant fields.
- Where possible, prefer documents that generate more than one candidate for each field. The system analyzes the neighboring words of candidates. It not only learns positive indicators for successful candidates, but also negative indicators.
- The number of samples needed to create a good extraction learnset depends on the task. In general, you need more samples to cover extraction than to cover classification.
- You can either learn all fields defined for a class at once, or one field at a time. In the first case, your project settings should support selection of the first field when switching to the next document, in the second case your project settings should make sure that the field selection is not changed when switching to the next document.

11.7.1 Create Learnsets for all Fields in a Class

To create learnsets for all fields in a class, complete the following steps:

1. Verify the following:

- Ensure that the **Execute Extraction** option is active on the **Runtime Mode** tab in the Settings dialog box.
 - Ensure that the **Add trained documents to the learnset** option is active on the **Train Mode** tab in the Settings dialog box.
 - Homogeneous sets of sample documents where all documents within a set belong to the same class, as learning data extraction performs class by class.
 - You can either learn all fields defined for a class at once, or one field at a time. In the first case, your project settings must support selection of the first field when switching to the next document, in the second case your project settings must make sure that the field selection is not changed when switching to the next document.
2. Switch to **Document Input Selection - Batch**.
 3. In the left pane, select the first document from the batch that contains the documents for your learnset.
 4. Switch to *Normal Train Mode*.
 5. In the lower left pane, on the **Classes** tab, double-click the class you want to train.

The Fields tab displays the fields of the selected class. The fields are empty and the first field is selected. Candidate highlighting turns on automatically.

Note: If you use field inheritance, select the parent class.

6. In the toolbar, click the **Classify / analyze current document** button to highlight all candidates for the first field on the document.
7. In the middle pane, click the correct candidate.

The field contains the text and the field's background color changes to green. The program automatically selects the next field and highlights its candidates. When all document fields have a value, the document is automatically added to the extraction learnset. The program continues automatically with the next document and the first field.
8. Repeat the previous steps until you have enough samples.
9. Optional. On the **Classes** tab, check the number of documents in the extraction learnset.

11.7.2 Create Learnsets for a Single Field

To create learnsets for a single field in a class, complete the following steps:

The prerequisites for this task is to make sure that the field selection is not changed when switching to the next document.

1. Switch to *Document Selection Mode*.
2. Select the first document from the batch that contains the documents for your learnset.
3. Switch to *Normal Train Mode*. On the left side of the window, the **Classes** tab is in the foreground.

4. In the list in the lower section of the **Classes** tab, double-click the class you want to train. In the **Fields** tab, the fields of the selected class display. The fields are empty and the first field is selected. *Candidate highlighting* is enabled automatically.
5. On the **Fields** tab, select the field.
6. On the toolbar, click **Classify/Analyze** <current document> to highlight all candidates for the selected field on the document.
7. To select the correct candidate, click it. The candidate's text is written to the field, and the field's background color changes to green.
8. In the toolbar, click **Add Document to Learnset** to add the current document to the extraction learnset. Designer continues automatically with the next document and the selected field.
9. Repeat the previous two steps until you have enough samples.

11.7.3 Improve Learnset

To improve the learnset, complete any of the following steps:

- Skipt Fields
- To repeat the training of a field, complete the following sub steps:
 1. In the left pane, on the **Fields** tab, select a field.
 2. On the toolbar, click the **Classify /analyze current document** button.
 3. To add the current document to the extraction learnset, on the toolbar, click the **Add document to Learnset** button. You can use a document even if some fields are empty.
- You can use words to fill fields that were not found as candidates. Complete the following sub steps:
 1. Click the **Highlight Words** button.
 2. In the left pane, on the **Fields** tab, select the required field.
 3. In the middle pane, click the word you want to add.
 4. In the **Add Candidate** dialog box, click **Yes**.
 5. In the right pane, on the **Analysis** tab, review the settings to ensure that the candidate can be found.

11.7.4 Modify Learnset

To modify the extraction learnset, complete the following steps:

1. Switch to *Document Input Selection Mode - Learnset*.
2. In the upper left pane, right-click a class and then click **View Extraction Learnset**.
3. In the lower left pane, add or remove documents from the learnset by selecting or clearing the check boxes.
4. Learn the project again to apply the changes.

11.8 About Learning

During the learning process, WebCenter Forms Recognition takes your learnsets and the manually assigned candidates. It identifies the correlation between successful candidates, unwanted candidates and their environment. It thus learns to extract data automatically if they are similar to the ones that you have selected.

Learning is required:

- After you set up your extraction scheme and created learnsets.
- When you have changed field definitions by adding, deleting, moving or renaming fields.
- When you have changed any parameters affecting the analysis and evaluation methods.
- When you have changed a learnset.

Note: Classification and extraction are usually learned together. To learn extraction, proceed as described for classification. For more information, refer to [Section 7.12.9: About Learning](#).

11.8.1 Check the Learn Status of a Field

You can check at any time whether a field has already been learned or whether it requires relearning. To check the learn status of a field, complete the following steps.

1. Switch to *Definition Mode*.
2. On the **Classes** tab on the left side of the window, double-click a class.
3. On the **Fields** tab, select a field.
4. Click the **Evaluation** tab on the right side of the window.
5. Click the **Brainware Extraction** tab. It displays the learn status.

11.9 Test the Extraction

To test the extraction, you must have defined and configured the analysis and evaluation engines. You cannot test the extraction in Train Mode, as the candidate selection is always done by the user in that mode.

You must have the following items to test the extraction:

- The classification must work.
- If *Brainware* extraction is required, the evaluation must have been trained.
- Ensure that the **Execute Extraction** option is active on the Runtime Mode tab in the Settings dialog box.

11.9.1 Test the Extraction in Definition Mode

To test the extraction Definition Mode, complete the following steps:



1. Open a document.
2. Complete one of the following sub steps:
 - To test the classification only, in the left pane, select the **Classes** tab.
 - To test classification and extraction, in the left pane, select the **Fields** tab.

3. Optional. Complete the following sub steps:
 - a. To highlight all candidates for the currently selected field in maroon, on the toolbar, click the **Highlight Candidates** button.
 - b. To view the weight of a candidate, in the middle pane, move the mouse over the candidate.
4. Complete the following substeps.
 - a. To highlight all fields in the document, on the toolbar, click the **Highlight Fields** button.




Note: Valid fields highlight in green, invalid fields highlight in red.

 - b. To view the name of a field as a tooltip, in the middle pane, move the mouse over the field.
5. To analyze the documents, complete any of the following sub steps:
 - To process the current document, on the toolbar, click the Classify / analyze current document button.
 - To process the next document, on the toolbar, click the **Process next document** button.
 - To process all documents in the current set starting with the current one, on the toolbar, click the **Process documents starting with current document** button.

Use one of the following highlighting options to determine candidates, weights and validity:

Button	Description	View Menu
	Highlights all candidates for the currently selected field in maroon. A candidate is a possible value for a field identified in the processing step. To view the weight of the candidate, point to it with the mouse. The weight displays as a tooltip.	Highlight Candidates
	Highlights all fields in the document. If the field is valid it is highlighted in green, otherwise it is highlighted in red. To view the name of the field, point to it with the mouse. The name displays as a tooltip.	Highlight Fields

Use one of the following buttons to analyze the documents. The test results display on the **Fields** tab on the left side of the window.

Button	Description
	Processes the current document. The button's drop-down menu enables or disables the debug mode.
	Processes the next document.
	Processes documents in the current set starting with the current one.

11.9.2 About Test Results

Possible results on the field level are:

The Field Is Empty	This happens if no candidate was found or no candidate had the required weight.
The Field Is Invalid	This is indicated by the red background color and happens if the best candidate had the required weight, but not the required distance or the best candidate did not meet standard or script-based validation requirements.
The Field Is Valid	This is indicated by the white background color.




11.9.3 Test the Extraction in Runtime Mode

To test the extraction in Runtime Mode, complete any of the following steps:

1. Open a document.
2. Switch to *Runtime Mode*.
3. To process all documents in the current set starting with the current one, on the toolbar, click the **Process documents starting with current document** button.
4. Optional. To pause the processing, on the toolbar, click the **Wait at current position** button.
5. Optional. To stop the processing and clear any results from previous runs, on the toolbar, click the **Stop / Clear** button.
6. In the upper left pane, on the **Statistic View** tab, observe the results.
7. Optional. In the upper left pane, on the **Statistic View** tab or the **Class Tree View** tab, select a class and observe the results in the lower left pane.
8. Optional. In the lower left pane, select an entry and observe the extraction and reference results in the lower right pane.

Note: The panel displays the field values determined for the current document and compares them with the results for the reference documents, where valid results are marked with a green check mark and invalid result are marked with a red cross. Differences between the test result and the reference result are indicated in red.

Use one of the following buttons on the toolbar to analyze the documents:

Button	Description
	Processes documents in the current set starting with the current one.
	Pauses the processing of documents.
	Stops the processing of documents. Clears any results from previous runs.

11.9.4 Extraction Results in Verifier Test Mode

If you use table analysis and want to check the extraction results, use the Verifier Test Mode. The test options described above only tell you whether a table was found at all, and whether the extracted result is valid. Which data was actually written to the table cells is only visible in Verifier Test Mode.

11.10 Extraction Optimization

Most extraction problems are caused by the following problems:

- Incomplete configuration.
- Insufficient quality of the learnset.
- Insufficient quality of the OCR.

11.10.1 Resolve Configuration Problems

An incomplete configuration most likely results in empty fields for all documents in a test set. The problem may be restricted to certain fields. To resolve configuration problems, complete any of the following steps:

- For a given field, use candidate highlighting to check whether your analysis settings deliver enough candidates.
- For a given field, check the candidates' weights.
- If the candidates' weights are all zero, check whether an evaluation method is specified at all. For more information, refer to [Section 9.11: Specify the Analysis Engine](#).
- Check whether the evaluation has been learned. For more information, refer to [Section 11.6: About Learning the Extraction](#).
- If you use field inheritance, check the parent and child configuration.

11.10.2 Resolve Learnset Problems

Problems with the learnset most likely result in low average extraction levels with empty and invalid fields. For some documents in the test set, the extraction works, for others not. The problem most likely affects all fields that use the learnset. To resolve learnset problems, complete any of the following steps:

- Check the candidates' weights and the distances. If your weights and distances are too low, you need more samples.
- Add candidates which were previously not classified correctly due to the initially insufficient learnset.
- Use candidates from the same document with an almost identical weight. These candidates are suitable to differentiate.
- Avoid candidates with a high weight. These candidates do not improve the learnset, because they would already be selected correctly.

11.10.3 Resolve OCR Problems

Problems occurring in the OCR processing makes the extracted data unusable. If you use a standard validation (No Rejects), you will obtain low average extraction levels with many invalid fields. To resolve OCR problems, complete any of the following steps:

- The problem may be caused by the quality of your images. Make sure that your document input is properly prepared. For example, check scanner settings or fax export settings.
- The default OCR settings may not be suitable to process your document input. If candidates are found with many unrecognized characters, change the OCR settings at the zone level or at the field level. If the OCR is so bad that you

cannot find candidates, you may have to change the OCR settings at the project level.

11.11 About Extraction Retaining

Use this feature to combine new extraction results with those already available in the workdoc if double or triple document processing is required.

The feature affects all programs that apply extraction processing in Oracle WebCenter Forms Recognition. To use the feature, either configure two separate projects or perform the secondary extraction using a script.

For example using two projects, refer to [Section 11.11.1: Extraction Retaining using Two Projects](#).

For example using a script, refer to [Section 11.11.2: Extraction Retaining Using a Script](#).

11.11.1 Extraction Retaining using Two Projects

Project *P1* contains the definition of a class hierarchy with the root *Header Fields* class. The *Retain Previous Extraction Results* feature is disabled. A Runtime Server instance applies an extraction workflow step with *P1* redirecting the documents to state **551** in both failure and success cases.

Another Runtime Server instance applies extraction with the second project *P2* that contains the definition of the *Table Fields* class. The *Retain Previous Extraction Results* feature is enabled for project *P2*. This Runtime Server instance has **551** as its input extraction state and **552** and **700** output workflow states for extraction failure and success respectively. The resulting workdocs contain all desired fields *A*, *B*, *C* and *T*.

11.11.2 Extraction Retaining using a Script

There is just one project *P1* that contains both class hierarchies; one with *Header Fields* and one with *Table Fields*. By default, it always applies extraction through the first hierarchy but, if required, can also launch the second level processing through a script like the example shown below.

```
Private Sub Document_PostExtract(pWorkdoc As SCBCdrPROJLib.SCBCdrWorkdoc)
    On Error GoTo ERROR_HANDLER
    Project.AllClasses.ItemByName("Table Fields").Extract pWorkdoc
ERROR_HANDLER:
    Project.LogScriptMessageEx CDRTYPEError, CDRSeveritySystemMonitoring,
    "Second level extraction has failed."
End Sub
```

11.11.3 Activate Extraction Retaining

If double or triple document processing is required, activate the extraction retaining. Complete the following steps:

1. On the **Options** menu, click **Settings**.
2. In the Settings dialog box, on the Processing tab, select the When processing document always retain previous extraction results and then click **OK**.

11.12 Text Field Candidate Evaluation

Text field candidate evaluation works like classification evaluation.

For example, you have a given text field, a given document, and Brainware extraction as an evaluation method that returns weights for each candidate.

Candidate evaluation uses two values defined at the field level to decide whether there is a valid candidate.

- A minimum weight called threshold (by default 50 percent)
- A minimum difference in weight between the best candidate and the second-best competitor called distance (by default 10 percent.)

Both requirements must be true to extract a valid candidate.

Consider the following example:

	Weight of Candidate A	Weight of Candidate C	Weight of Candidate C
Field A	0.45	0.36	0.28
Field B	0.30	1.00	0.00
Field C	0.55	0.49	n/a

11.12.1.1 Result

For *Field A*, there is no valid candidate. The weight of *Candidate A* is below the threshold, and the distance to *Candidate B* is below the minimum distance.

For *Field B*, *Candidate B* is a valid candidate. Its weight is above the threshold. In addition, the distance to the second-best competitor, *Candidate A*, is above the minimum distance.

For *Field C*, there is no valid candidate. The weight of *Candidate A* is above the threshold, but the distance to *Candidate B* is below the minimum distance.

11.13 Table Field Candidate Evaluation

In table analysis, each candidate is evaluated using the criteria listed below:

Table Property	Range of Results
Fraction of default columns identified.	0.0 - 1.0
Fraction of table columns mapped.	0.0 - 1.0
Fraction of entries found in cells which require entries.	0.0 - 1.0
Fraction of neighboring cells without overlap (column view).	0.0 - 1.0
Fraction of neighboring cells without overlap (row view).	0.0 - 1.0
Fraction of correct formats.	0.0 - 1.0

For each table candidate, the total significance is determined as the average of the results per property. All table candidates, which reach a certain threshold for the total significance, are transferred to the workdoc, ordered by the total significance. By default, the candidate with the highest significance is used for data extraction.

Note: During table analysis, field-level validation settings defined through the user interface are ignored. However, you can modify the evaluation settings using scripts.

12 Setup Supervised Learning

The Supervised Learning Workflow (SLW) is used to automatically create and learn new document classes. This automatic process creates the new derived document classes for a base document class. This document class is not inherited from any other document class.

The derived document classes cannot have further derived document classes.

You must learn any documents that you add to the local learnset before they are released. Otherwise, the subclass will not be created in the project. In addition, the local project must be saved in order for the subclass to be created in the global project.

12.1 Considerations and Constraints for Supervised Learning Workflow

The Supervised Learning Workflow automatically creates and learns new classes as subclasses of a base document class. A base document class represents the highest level of classification and is not inherited from any other class.

- These automatically created and trained classes are always placed directly under the base document class to which the document used for automatic training is currently classified.
- The classification field property that creates a document class must not be empty.
- Assign the Associative Search Engine as the analysis engine for the classification field.
- The Associative Search Engine settings for the classification field must include a definition of the document class format. Define this format in a way that it ensures a unique name for every document class that could possibly be created. One way to ensure that the names are unique is to use the primary key or any other unique field from the database for the vendor pool as part of the class name.
- Enable the Supervised Learning Workflow in Verifier and Verifier.
- You must manually configure the classification for base document classes. Manual configuration can consist of automatic redirection to a base class, or through any available classification engine.

12.2 Apply the Supervised Learning Workflow

To apply the Supervised Learning Workflow to an existing project, complete the following steps:

1. Ensure that RTS is configured and running.
2. In **Verifier**, open the project.
3. On the **Options** menu, click **Settings**.

4. In the **Settings** dialog box, on the **VerifierTrain Mode** tab, select the **Enable automatic Supervised Learning workflow for this project** and then click **OK**.
5. Save the project, close and reopen it.
6. Switch to *Definition Mode*.
7. To process the current document, on the toolbar, click the **Classify /analyze current document** button.
8. Switch to *VerifierTest Mode*.
9. Test the project-related validation.
10. In Oracle WebCenter Forms Recognition Verifier, activate the Supervised Learning Workflow.

Note: For more information, refer to *Configure Supervised Learning* in the *Oracle WebCenter Forms Recognition Verifier User's Guide*.

12.3 Example for a Supervised Learning Workflow

You have an existing vendor class, named Oracle Invoice that process invoices from supplier Oracle Corporation. You have a document class with the same purpose, but with a different name (such as, **ORCL1234567**) that can be created through the Supervised Learning Workflow, but may cause conflicts.

To avoid conflicts, it is necessary to take into account the following cases to combine non-SLW and SLW processed class nodes:

- **Case A.** The existing set of classes does not coincide with the set of classes that the Supervised Learning Workflow might create.
- **Case B.** The existing set of classes does coincide with the set of classes that the Supervised Learning Workflow might create. Automatic learning for these classes is not required.
- **Case C.** The existing set of classes does coincide with the set of classes that the Supervised Learning Workflow might create. Automatic learning for these classes is required. In addition, these classes are Level 1 classes, meaning that they are derived from a base document class.
- **Case D.** The existing classes do coincide with the set of classes that might be created by the Supervised Learning Workflow. Automatic learning for these classes is required. These classes are not on Level 1 in **WebCenter Forms Recognition** project classes' hierarchy or their base classes do not satisfy constraints for Supervised Learning, meaning that *Classification Field* settings are not correctly configured.

Criteria	Case A	Case B	Case C	Case D
Existing classes conflict with potential classes.	No	Yes	Yes	Yes
Automatic Learning required.	n/a	No	Yes	Yes
Level 1 class (derived from base class).	n/a	No	Yes	No
Base class do not satisfy requirements for SLW execution or classification field improperly configured.	n/a	n/a	n/a	Yes

Note: Classes created by the Supervised Learning Workflow cannot have the same names as existing document classes. To ensure that this never happens, you must setup the *Class Name Format* in the Associative Search Engine settings in a unique way, so that existing class names and new class names never overlap. For more details about Supervised Learning Workflow constraints, refer to [Section 6.1.13.6: Planning Supervised Learning](#).

Case A

This case is the easiest to handle. You only need to create new base classes or use existing base classes required for the Supervised Learning processing, and then run the Supervised Learning Workflow in conjunction with existing classification and extraction project configurations without any additional adjustments.

Case B

Case B occurs when you want to keep an existing document class with its learnset, scripting, and other setting, as is, and all documents should still be classified to this class. It is conceivable that Supervised Learning could create a new document class by automatic learning for this document. In that case, you must disable automatic learning for this document class.

To disable automatic class creation for a particular class, set the *Vendor_Type* field in the CSV file to **2** for the corresponding class search entry. For details, refer to [Section 9.17: About Associative Search Engine](#).

Note: If the *Vendor_Type* column is not part of the vendor pool, you must create or designate a new column in the vendor pool CSV file to be used for Associative Search Engine search database import. This field can have a value of **0**, **1**, or **2**. By default, when the field is not available in the vendor pool, the system acts as if the *Vendor_Type* value is **0**. A *Vendor_Type* value of **0** means that all documents with more than n% invalid fields are added to the learnset. For more details, refer to the *Oracle WebCenter Forms Recognition Verifier User's Guide*.

You must re-launch the pool import in the Associative Search settings to establish the *Classification Field* column as the *Vendor_Type* field. Do not use this field as a search field.

In the scenario, there is an entry **Oracle** in the vendor pool, but learning is prohibited for this entry. Although no new document class for Oracle will be created, you still must ensure that the system classifies the documents to the 'old' document class, **Oracle Invoices**, and not to **Invoices**. For the example, those document classes are situated within different tree branches. To ensure correct classification, the following project configurations are possible. There are other possible configurations, depending on the specific project being migrated.

- Move part of the document class tree to the new **Invoices** class. Automatic learning for **Oracle Invoices** is disabled and all the configurations (such as for the learnset) are retained. Because this part of the class hierarchy does not have to be trained, the derived document class does not have to at Level 1.
- Do not change the document class hierarchy. For the document class **Invoices** do not select a classification, but do select one for the **Companies** branch. Additionally, set a default classification to **Invoices**. This means that if a document cannot be classified to the document class for which classification engines are selected, it will be classified to **Invoices**.

- Use WinWrap Basic scripting to ensure that the document is classified to the correct document class.

Case C

If Supervised Learning must be enabled for an existing class that is a derived document class at level 1, and its base class satisfies constraints for Supervised Learning execution, *Vendor_Type* in the CSV file must be **0** for the corresponding class entry in this file.

In addition, the existing class name and the class name automatically generated by Supervised Learning Workflow and specified on the Associative Search Engine settings for *Class Name Format* must coincide with each other. To do this, either:

- Rename the existing class so that its name coincides with the name that is generated through the *Class Name Format* for the corresponding entry in a CSV file.
- Change *Class Name Format* so that the names are the same. This can only be applied if it can be done for all classes that *Case C* covers. In addition, uniqueness of all new document class names must be guaranteed.
- Create a new (non-search) field in the CSV file and give this field a descriptive name. In the CSV file generation procedure for all classes described by *Case C*, use their class names to fill out the field. For all other classes, create a name according to field naming conventions. Then, in the settings for the Associative Search Engine, specify the new column name as the *Class Name Format*.

The base document classes must satisfy the constraints for execution of the Supervised Learning Workflow, i.e. the *Classification Field* settings are configured correctly, the generic document class is at level 0 of the classification hierarchy and the Associative Search Engine is configured. For details, refer to [Section 6.1.13.6: Planning Supervised Learning](#).

Case D

If the Supervised Learning Workflow must be enabled for a particular existing class that is not a Level 1 class, or its base class does not satisfy constraints for SLW execution, the *Vendor_Type* field value must be **0** for the corresponding class entry in this file.

In this case, follow the instructions covered for *Case C*.

In addition, the *Case D* class must be moved under another base class to ensure that the document class is now at Level 1. The easiest way to do this in Designer is to switch to the **Classes** view in *Definition Mode* and use drag & drop to move it to the required class. You can also implement a scripted procedure, use *Project.MoveDocClass*, and save the project.

12.4 Train the Base Class

A base document class – which is also called base DocClass – is a generic category that applies to all processed documents, such as “Invoices”. You must pre-configure the classification for this class, as it is not trained automatically, by assigning a classification engine or assigning the base document class as the default classification result. Complete the following steps:

- To select a classification engine, complete the following sub steps:
 1. Switch to *Definition Mode*.

2. In the left pane, on the **Classes** tab, right-click the base document class and then click **Show Properties**.
 3. In the right pane, on the **Classification** tab, select an engine.
- To define a class as default classification result, complete the following sub steps:
 1. Switch to *Definition Mode*.
 2. In the left pane, on the **Classes** tab, right-click the project and then click **Show Properties**.
 3. In the right pane, on the **Classification** tab, from the **Default Classification Result** list, select a class.

Note: The Supervised Learning Workflow can only be applied for base classes and their direct sub-classes. Only administrators can select base classes that become SLW nodes, where SLW node is a class node that defines how its sub-nodes are automatically created and learned. The simplest way to define a project's SLW node classification is through its default classification result. The method works if there is only one SLW node.

4. Insert at least one header field (usually called a text field) for which the **Associative Search Engine** is selected as the analysis engine.
5. Set the Classname Format on the Analysis Editor. The header field is the one that is used to find the document class name used to create a new derived document class (if the class does not already exist), and then to automatically classify the document.
6. Additionally, the Classification field on document class-level must be defined. Select the Document Class tab on the right side of the screen, and notice the drop-down box for Classification field. Select your generic document class for this setting from the drop-down list.

12.5 Establish Multiple Classification Results using a Script

This technique enables you to keep the generic extraction pointed towards the default class, while moving the validation script to a different class. This can be configured and tuned manually through the various classification engines. To establish multiple classification results using a script, complete the following steps:

1. Define a class as default classification result. For details, refer to [Section 12.4: Train the Base Class](#).
2. Create a script that executes in the *Document_PostExtract* event and use the method *pWorkdoc.DocClassName* to classify the document to a different class.

12.6 Training Other Fields

For all other text fields that are defined for a generic document class, besides the one for which the Associative Search Engine is selected, you must use the Format Analysis Engine for analysis and the Brainware Extraction Engine for the evaluation engine.

If you have table fields, you must use Brainware Table Extraction for those fields.

12.7 Create Derived Document Classes

You do not have to manually configure the derived document classes; they will be created automatically when the documents are added to the learnset.

When the documents are added to the learnset, a new derived document class is automatically created. The new document class name corresponds to the settings established for the document-level settings of the field **Class Name Format** on the *Analysis Editor*. Therefore, this must be defined in a way that ensures uniqueness of every document class name that could be created, that means that the ID, or any other unique field, should be part of the **Classname Format**.

Though not obligatory, it can be useful to train some additional fields for the generic class now. This would mean that there would already be extraction results for the new class, and the Verification form would therefore already be partially populated.

To train the generic document class now, switch to *Train Mode*. If there is only one generic document class, you can set it as the default classification result. When this option is set, all documents are classified to the generic document class and only the extraction results have to be checked.

12.8 Create Derived Document Classes with the Supervised Learning Workflow

To create the derived document classes, complete the following steps:

1. Switch to *Verifier Train Mode*.
2. Check the extraction results.
3. Add the documents to the learnset.

12.9 About the Supervised Learning Workflow Process

For the new automatically learned classes the system uses fixed and pre-determined engines that cannot be adjusted. Even though a new learned *DocClass* has pre-defined engines that cannot be adjusted, the classes used in Supervised Learning can have any other engines assigned. When new classes are created automatically, they initially have the configurations of the engines mentioned below. However, the configuration can be changed, adjusted, or tuned manually.

The Supervised Learning Workflow uses the following engines:

- *Brainware Layout Classification* to automatically train level 1 classification usually called vendor-classification in the context of PIC projects used for invoice processing.
- *Associative Search Engine* to automatically detect a new class name, and, optionally, to automatically train level 1 classification.
- *Brainware Extraction* to train automatic extraction of header fields for fields that have *Brainware Extraction* and *Format Analysis* engines assigned for the parent base document class.
- *Brainware Table Extraction* to train automatic extraction of header fields for fields that have *Brainware Table Extraction* engine assigned for the parent base document class.

12.10 Adding Supervised Learning to Brainware Layout Classification

The Brainware Layout Classification engine is used to automatically train the derived (Level 1) document classes. However, the user can also choose that the results of the Associative Search engine will be taken into account for the classification. These settings are established on the Supervised Learning Editor.

In *Definition Mode*, complete the following steps:

1. Select the project node in the **Classes** tree.
2. Select **Show Properties** from the **Edit** menu.
3. Select the **Supervised Learning** tab.
4. On this tab, you can choose from **Make supplier field invalid** and one of the following options:

Do Smart Decision	The system determines which one is the right <i>DocClass</i> , based on an algorithm that compares the results of the associative search and the <i>Brainware Layout Classification</i> .
Assign Current Document Class Name to Supplier Field	The system determines which document class is used to assign the document to the class name in the field you created, based on the result of the <i>Brainware Layout Classification</i> .
Assign Supplier Field Content to Current Document Class Name	The system uses the result of the <i>Associative Search Engine</i> if the document class is available, otherwise the result of the <i>Brainware Layout Classification</i> will be used.
Do Nothing	The system uses only the result of the <i>Brainware Layout Classification</i> and not the result of the <i>Associative Search Engine</i> .
Allow Interactive Table Extraction for Verifier User	Activates the Correct Tables button in Verifier.
In Verifier Apply Local Extraction Only for Invalid Fields	Local classification and extraction are carried out only for invalid fields, either manually through the Analyze Document button or automatically through settings.

12.11 Specify the Default Classification Result

To specify the default classification result, complete the following steps:

1. In the left pane, on the **Fields** tab, insert any additional fields that need to be learned.
2. In the left pane, on the **Classes** tab, select the project.
3. In the right pane, on the **Classification** tab, from the **Default Classification Result** list, select your Base DocClass.

Note: Oracle WebCenter Forms Recognition assigns all documents that the Associative Search Engine cannot classify to this class.

4. On the **Supervised Learning** tab, select **Do Smart Decision**.

12.12 ASE Pool Configuration

The following fields are mandatory:

Vendor ID	It is a unique sequential designation for each record in the file.
Internal_ID	It must be a unique numerical designation for the vendor name.
Vendor_Name	Is a short form of the vendor's full name.
Vendor	Is the full name of the vendor.

The following field is optional:

Vendor_Type	Used to influence a Supervised Learning setting that indicates the threshold for invalid fields, that, if exceeded, forces a document, or, by extension, members of a document class, to be learned. In most cases, this value is set to 40 percent. You can always another field as vendor type, as long as the values are correct. 0: Adds all documents classified by that vendor name to the local learnset. This is the default value if no <i>Vendor_Type</i> field is available in the CSV file. 1: Adds only the first document in the class to the local learnset. 2: Adds none of the documents in the class to the local learnset.
--------------------	---

Ensure that field names do not contain periods or other special characters.

You can optionally add header rows. The name of the header columns can differ, but the content must match the data expected for the fields mentioned above.

12.13 Options for Creating Derived Classes in Verifier

In Verifier, you can set the **Automatically if more than N% invalid fields with Always prompt before adding** option on the Self-Learning tab to define that a document is automatically added to the learnset when a certain percentage of the fields are invalid.

You can also set the **Always prompt before adding** option to be prompted whenever a document is added to the learnset.

For more information, refer to *Supervised Learning Settings* in the *Oracle WebCenter Forms Recognition Verifier User's Guide*.

13 Verification

13.1 About Quality Assurance with Verifier

The quality assurance program Verifier provides a way to check and correct the results of automatic document classification and data extraction. In addition, it enables you to manually index documents.

Verifier users are in charge of validating the documents, which are presented as batches. Whether a batch is valid or not determined using the following rules:

- A batch is valid if all of its documents are valid.
- A document is valid if it has a valid classification result and if it has a valid extraction or indexing result.
- An extraction result is valid if all the fields are filled with valid content.

The Verifier user is presented with all invalid or uncertain results from a given batch and needs to correct and confirm them.

Normally, results that are invalid by definition cannot be confirmed by the user, but in the project definition, you can allow the user to validate invalid field results in exceptional cases. This procedure is called *forced validation*.

To enable you to adjust Verifier to the requirements of your project, two modes are available in Designer:

- The *Verifier Design Mode*, in which you set up forms to use in the verification step.
- The *Verifier Test Mode*, in which you can dynamically simulate verification.

The Verifier user interface for classification verification can use the same form all the time, so nothing needs to be defined there. What you need to create are the forms that will be used for verification of data extraction or for manual indexing. This step can be simplified by using default forms, but it can also be a complex task if there are special requirements. In addition to the features offered through the user interface, you can implement custom verification routines in WinWrap Basic scripts.

In WebCenter Forms Recognition, verification steps take place when a batch of documents reaches a predefined state that is represented by an integer number. The state changes again when verification is complete. States are used to distinguish valid and invalid results. For more information about input and output states, consult these manuals:

- *Oracle WebCenter Forms Recognition Runtime Server User's Guide*: The settings of this module determine the input and output states in the entire WebCenter Forms Recognition workflow.
- *Oracle WebCenter Forms Recognition Verifier User's Guide*: The settings of this module determine the input and output states of verification steps.

13.2 About Verification Projects

Designer organizes verification forms in verification projects. It uses the following structure:

- Each WebCenter Forms Recognition Designer project can contain one verification project.
- Each verification project contains a set of forms that are assigned to document classes.
- Each document class has one form that will be used by default. The remaining forms are only used when a document that requires validation is in a certain pre-defined state.
- Each verification has up to four areas:
 1. Form field area.
 2. Current input area.
 3. Document display area.
 4. User information area.

13.2.1 Configure Project Validation Properties

Prerequisites: Classes and fields are already defined.

To set the project validation properties, complete the following steps:

1. Switch to *Verifier Design Mode*.
2. In the class selection panel at the upper left edge of the window, right-click the root node that represents the project.
3. From the popup menu, select **Project Properties**. The *Validation Settings* dialog displays.
4. Under **Force Validation**, select whether Verifier users can validate a field, even if its content is invalid by definition:

Permitted	The user can force validation by pressing the [Enter] key three times.
------------------	--

Forbidden	The user cannot force validation.
------------------	-----------------------------------

5. Under **Designer Test Mode Document Routing**, if you have script code associated with the *ScriptModule_RouteDocument* event, you can use the states entered in this area to test your scripting in Designer's *Verifier Test Mode*. These states are not used in Verifier.

13.2.2 About Creating Verification Forms

Designer provides the following ways to create a verification form:

- You can create empty forms first and add the form elements then.
- You can create default forms that automatically contain a form field, a label, and a viewer for each field that will be visible in Verifier.
- You can create forms for all classes in one step.

- You can change fonts, colors, and background colors for verification forms and elements. For more information, refer to *SCBCdrVerificationForm* in the *Oracle WebCenter Forms Recognition Scripting User's Guide*.

13.2.2.1 Verifier Design Mode User Interface







The Verifier Design Mode user interface provides the following panes:




















Classes pane	The upper left pane displays a hierarchical representation of the classes in the current project.
Forms pane	The middle left pane displays the name and properties of the forms.
Fields pane	The lower left pane displays a table with the mapping of data extraction fields and verification form fields of the current form
Form pane	<p>The middle pane displays the controls, such as form fields, labels, buttons, viewers and tables that you use in Verifier for correction and manual indexing.</p> <p>Each control must be assigned to a data extraction field. The viewers display the document area that Verifier extracted to fill a given field.</p>
Document pane	The right pane displays the current document.
Current Input pane	The lower right pane displays a magnification of the current form field and the associated viewer. You can use it for convenient viewing and editing.
User info pane	This pane works like a status bar. You can display custom messages here that explain why a field is invalid and provide hints how to correct invalid results.

Note: The position of the form areas may vary according to the selected layout, or only a subset of form areas may be present.

13.2.2.2 Verifier Design Mode Toolbar

Use the following buttons to control form and document display during the verification procedure:

Button	Description
	Insert default verification forms for all doc classes
	Delete all verification forms
	Insert default form/Insert empty form
	Delete verification form
	Maximize Verification form view
	Show/hide script

	Show selected object properties
	Show/hide Grid
	Enlarge Grid Step
	Shrink Grid Step
	Data View Layout Style 1 The Verifier window displays the panes as follows. <ul style="list-style-type: none"> ▪ Document in the upper pane ▪ Verification form in the middle pane ▪ Magnification view of the current field in the lower pane
	Data View Layout Style 2 The Verifier window displays the panes as follows. <ul style="list-style-type: none"> ▪ Verification form in the middle left pane ▪ Document view in the middle right pane ▪ Magnification view of the current field in the lower pane
	Data View Layout Style 3 Similar to layout style 2, but without the magnification view.
	Data View Layout Style 4 Similar to layout style 1, but without the magnification view.
	Data View Layout Style 5 Preferred layout style for projects with 1 to 5 fields to verify. The Verifier window displays the panes as follows. <ul style="list-style-type: none"> ▪ Document view in the upper pane ▪ Magnification view of the current field in the lower pane
	Data View Layout Style 6 The Verifier window displays the panes as follows. <ul style="list-style-type: none"> ▪ Document view in the left pane ▪ Verification form in the right pane ▪ Magnification view of the current field in the lower pane
	Data View Layout Style 7 Similar to layout style 6, but without the magnification view.
	Add New Label Element
	Add New Viewer Element
	Add New Button Element
	Delete Selected Elements
	Define Smart Indexing
	Delete Smart Index
	Show Action List
	Lock Verifier controls

13.2.2.3 Create an Empty Form

To create a single empty form, complete the following steps:

1. In the class selection panel at the upper-left edge of the window, select a document class.
2. Complete one of the following actions:
 - From the menu, select **Edit ▾ DocClass ▾ Insert Empty Form**.
 - From the popup menu of the document class, select **Insert Empty Form**.

- Click the drop-down arrow associated with the **Insert New Default Verification Form** button and select **Insert Empty Form**.

13.2.2.4 Create Default Forms

To create a single default form, complete the following steps:

1. In the class selection panel at the upper left edge of the window, select a document class.
2. Complete one of the following options:
 - From the menu, select **Edit ▾ DocClass ▾ Insert Default Form**.
 - On the shortcut menu of the document class, select **Insert Default Form**.
 - Click the drop-down arrow next to the **New Default Verification Form** toolbar button and select **Insert Default Form**.

13.2.2.5 Create a Default Form for Each Class in the Project

To create a new verification project with one default form per document class, complete one of the following options:

- From the menu, select **Edit ▾ Project ▾ Insert Default Forms**.
- In the class selection panel at the upper-left edge of the window, right-click the project node, then from the popup menu, select **New Default Forms**.
- In the toolbar, click the **Insert New Default Verification Forms for All DocClass** button.

Note: If you create a new project, an existing project including its forms will be deleted when you confirm.

13.2.2.6 Configure Form Validation Properties

To set the form validation properties, complete the following steps:

1. Switch to *Verifier Design Mode*.
2. Display a form and make sure that no form elements are selected.
3. Complete one of the following options:
 - In the vertical toolbar, at the right edge of the window, click the **Show Selected Object Properties** button.
 - Right-click the form's background and select **Properties** from the popup menu.
4. In the **Properties** dialog box, enter the following information:

Name	Displays the name of the form. This field is read-only.
Assigned DocClass	Displays the name of the associated document class. This field is read-only.

Process Step Filter To use this form only for documents with a given state, select one of the pre-defined states from the list, or enter a value into the field. For default forms, keep the initial value **-1**.

Default Form Select this box to use the form as the default form, such as in all verification steps without a special predefined input state.

Document Validation Mode Select **Complete Validation Required** if only one user verifies an entire batch or **Partial Validation Allowed** if several users validate a batch in stages. If complete validation is required, the Verifier user must validate all the fields in the documents of a batch before being able to release the batch. If partial validation is allowed, only the fields on the form must be validated. With partial validation, you can employ multiple forms, i.e. one for each stage. You can set an output state for partially validated documents in Designer, which overrides the output state set in Verifier.

5. Click **Validation Settings** button, modify the settings if required.

Note: For more information about the *Validation Editor*, refer to [Section 10: Validation](#).

13.2.2.7 Display Forms

To display a form, complete one of the following steps:

- In the upper left pane, select a class to display its first form.
- In the middle left pane, select a form from the forms table.

13.2.2.8 Display all Forms of the Project or a Class

To display all forms of the project or a class, complete one the following step.

Display all forms of the project On the **Edit** menu, click **Verification Form** ▾ **Show All**

Display all forms of a class In the upper left pane, select a class.

The table of forms contains the following columns:

Form Name	Displays the name of the form. It is generated automatically and consists of the prefix Form , the name of the associated document class, and a consecutive number.
------------------	--

Default	Displays Yes if the form is the default form of a document class, and nothing if this is not the case. Default forms are used for each input state, unless another form has been explicitly assigned to an input state.
In Filter	Displays the input state assigned to a form, or -1 if no special input state has been selected.
Out Filter	Displays the output state assigned to a form, or -1 if no special output state has been selected. Output states are relevant if partial validation of documents is allowed and useful in the context of exception handling.
Assigned To	Displays the name of the associated document class.

13.2.2.9 Delete Forms

To delete a single form, complete the following steps:

1. In the form selection panel on the left side of the window, select a form.
2. Complete one of the following actions:
 - From the menu, select **Edit ▾ Verification Form ▾ Delete Form**.
 - From the shortcut menu of the form, select **Delete Form**.
 - In the toolbar, click the **Delete Verification Form** button.

13.2.2.10 Delete All Forms

To delete all forms associated with a project, complete one of the following actions:

- From the menu, select **Edit ▾ Project ▾ Delete All Forms**.
- In the class selection panel, at the upper-left edge of the window, right-click the project node and select **Delete All Forms** from the popup menu.
- In the toolbar, click the **Delete All Verification Forms** button.

13.2.2.11 Form Elements

You can add the following elements to a form:

Forms Fields	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	A toggle selection affecting data input applied when you have one of two values to choose from, such as on/off or yes/no. Check boxes are derived from form fields. You can set it up the caption with the text desired and select the default view.
List	Contains values to select from. Depending on the settings, a list field can be used as a combo box. Used during manual verification, this selection works with automatic completion.

Labels	Optional captions that help users to identify form fields, viewers and tables.
Viewers	Snippets of document areas, normally those that were extracted to fill fields or tables.
Buttons	Fire actions for script events.
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.

13.2.2.12 Create and Modify Form Fields

13.2.2.12.1 Add Form Fields

To add a form field to a form, complete one of the following options:

1. From the lower left pane, drag a field to the target position on the form.
2. Optional. To change the properties, right-click the form field and then click **Properties**.
3. In the Properties dialog box, under **Size & Position**, modify the values to position the element with greater precision.
4. Optional. To show static and dynamic complementary information to aid in manual verification, select **Read Only**.
5. Optional. To auto-complete the word when you start typing, select **Auto-completion**.
6. Optional. To control which field is selected next when the user presses the Tab key, modify the value in the **Tab index** field.

Note: If the specified index is already in use by another verification field, its tab index and, when required, subsequent fields, increments by one automatically.

7. Optional. To apply a horizontal scroll bar, select the **Horizontal Scroll Bar** check box.
8. Optional. To apply a vertical scroll bar, select the **Vertical Scroll Bar** check box.
9. Optional. To modify the validation settings, click the **Validation settings** button.
10. In the Validation Settings dialog box, modify the settings if required.

13.2.2.12.2 Create Labels

To add a label to a form, complete the following steps:

1. In the vertical toolbar on the right side of the window, click the **Add New Label Element** button. The mouse pointer changes.
2. Click inside the form and drag the frame to the desired position and size. When you release the mouse button the *Change Caption* dialog is displayed.

3. Enter a caption for the label and click **OK**.
 - To change the caption, right-click on the label element and select **Change Caption** from the popup menu.
 - To delete a label from the form, select the label element then click the **Delete Selected Elements** button on the vertical toolbar.

13.2.2.12.3 Create Viewers

By default, a viewer displays one data field.

A viewer can also display several fields, depending on the field selection at runtime.

For different views of the same zone on a document, you can also assign the same field to several viewers.

What you see in the viewer depends not only on the viewer properties, but also on its dimensions. The viewer always covers the area specified in the properties, but actually shows a larger snippet of the page that matches the viewer dimensions. To obtain the best results, ensure that the viewer dimensions closely resemble the geometry of the area to display.

To add a viewer to a form, complete the following steps:

1. On the right vertical toolbar, click the **Add New Viewer** element button.
2. On the form, click and drag to create the viewer.
3. Right-click the viewer and then click **Properties**.
4. In the Properties dialog box, from the **Assigned Field** list, select a field.

Note: If the content that displays depends on the selected document only or on the field that is selected by the user at runtime, leave the field empty.

5. Optional. If you want the viewer to cover the entire page of the current document, select **Whole Image** and specify the page number in the **Page** field.
6. Optional. If you want the viewer to cover a fixed area of a page of the current document, complete the following sub steps:
 - a. Select **Fixed Zone** and specify the page number in the **Page** field.
 - b. To specify the distance of the viewer area from the top and left edge of the page, and the width and height of the area, modify the **Left**, **Top**, **Width**, and **Height** percentage values.
7. Optional. If you want the viewer to cover the area of the document that contains the words used to fill the associated field, complete the following sub steps:

Note: This works no matter whether the field assignment is set at design time or at runtime.

- a. Select **Field**.
- b. To specify that a white border displays around the area covered by the viewer, under **Border**, specify the **Left/Right** and **Top/Bottom** values.

- Optional. If you want the viewer to cover a rectangle around the area of the document that contains the words used to fill the associated field, select **Field in Zone** and modify the **Width** and **Height** values.

Note: This works no matter whether the field assignment is set at design time or at runtime.

- Optional. If you want the viewer to cover the lines around the area of the document that contains the words used to fill the associated field, select **Field in Line**.

Note: This works no matter whether the field assignment was set at design time or at runtime.

- Optional. Under **Size & Position**, modify the values for size and position of the viewer on the form.

13.2.2.12.4 Create Buttons

To add a button to a form, complete the following steps:

- To create a new button, in the vertical toolbar on the right side of the window, click the **Add New Button** button. The mouse pointer changes.
- Click inside the form and drag the frame to the desired position and size.
- To name the button, right-click the button and select **Change Caption** from the popup menu.
- In the **Change Caption** dialog box, enter a name in the box and click **OK**. The *Change Caption* dialog is displayed.
- Enter a caption for the label and click **OK**. For information about configuring actions triggered by buttons, refer to [Section 13.2.2.12.5: Define an Action](#).

13.2.2.12.5 Define an Action

Actions are triggers for scripted events that you can create to perform project specific tasks. For example, you can trigger any allowed modification of Oracle WebCenter Forms Recognition related objects, or you can set an action to open a program or invoke third-party COM interface methods.

For more information, refer to *OnAction* in the *Oracle WebCenter Forms Recognition Scripting User's Guide*.

To define an action, complete the following steps:

- On the **Edit** menu, click **Define actions**.
- To add an action, in the Define Actions dialog box, click **Insert**.
- Optional. Click the **Action** field and type a name for the action.
- Optional. Click the **Description** field and type a description for the action.
- Click the **Button** field and select the button that you want to link to the action from the list.
- Optional. To assign a shortcut to the action, click the **Accelerator** field.
- In the Accelerator Properties dialog box, select **Enable This Accelerator**.
- From the **Key** list, select a key.

- Optional. Select [Ctrl] and/or [Shift] key.

Note: Many key combinations are assigned to existing Verifier functions or menu options, or may be assigned as system shortcuts. Ensure that the selected key combination is not yet in use. Within Oracle WebCenter Forms Recognition, the combination [Ctrl] + [Shift] key is usually available to combine with any other keys.

- Click **OK** and then click **Close**.
- To handle the action in the *Document_OnAction* event in the script, on the toolbar, click the **Show / Hide Script** button.
- Modify the script according to your needs.
- Switch to **Verifier Test Mode** and test the action.

13.2.2.12.6 Delete an element from a form

To delete an element, complete the following steps:

- Select the element on the form.
- On the right vertical toolbar, click the **Delete selected elements** button.

13.2.2.13 Configure the Table Validation Properties

You can only have tables in a form if table extraction is configured for the corresponding document class. A form can have more than one table. When table extraction is configured, you must set the table validation properties.

If required, you can make your columns invisible. The invisible columns must be valid to validate the entire table.

In table validation, the single table cells behave like fields in regular data extraction. To configure the table validation properties, complete the following steps:

- On the form, right-click the table and then click **Properties**.
- In the Properties dialog box, complete one of the following actions.

Optimize the column width automatically according to the column content

- Select **Automatic Column Sizing**.
- In the **Maximum Column Width** field, type the value in pixels.

Specify a fixed width for the columns

- Select **Fixed Column Sizing**.
- In the **Column Width field**, type the value in pixels.

Enable the user to modify the column width at runtime

Select **Manual Column Sizing**.

- Optional. If you want the user to navigate through the table line by line from left to right, select **Validate Table Line by Line**.
- Optional. If you want the user to navigate through the table column by column from top to bottom, select **Validate Table Column by Column**.
- Optional. Under **Size & Position**, modify the values for size and position of the viewer on the form.

6. Optional. To lock the table against edits, select **Read Only table**.
7. Optional. To enable faster verification, select **Auto-completion**.

Note: Auto completion might include check boxes or combo boxes, as well as automatic word completion when typing text.

8. Optional. Select **Cell Position Check**.
9. Optional. To modify the column settings, click the **Column Settings** button.
10. Optional. In the **Table Columns Properties** dialog box, complete any of the following sub steps:
 - Use the arrow buttons to reorder the columns.
 - In the **Display Name** field, type a name.
 - To protect a column from edits, select **Read-only**.

13.2.2.14 Auto-Completion

Auto-completion helps to speed up data entry. When you start to type, auto-text completes the word, suggesting the best matching items 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 auto-selected text is also highlighted in the original document. To override auto-completion, continue typing the desired text.

The auto-completion feature for a header field automatically selects the best candidate from the available ones if Verifier is in *Highlight Candidates* mode. However, the viewer will be updated only if the candidate appears once in the document; otherwise the viewer will be blank when auto-text completes the word for the field.

Do not insert formatted text for auto-completion. Auto-completion does not work on formatted text and characters incorrectly read by OCR.

13.2.2.15 Configure Smart Indexing

Smart indexing is a method to fill fields using entries from a database table. It can be used in the automatic extraction step as well as during verification and manual indexing. In smart indexing, two types of fields are distinguished:

Smart Index Field The information related to this field must be available both in the documents that are to be processed, and in the database table. Often, a unique identifier such as **Invoice Number** or **Customer Number** is used for this purpose.

Index Lookup Fields The information related to these fields must only be available in the database table, but not in the documents. This is because these fields are filled using a database lookup with the value of the smart index field as parameter. The lookup returns the table rows that will be used to fill the lookup fields.

13.2.2.15.1 About Database Support for Smart Indexing

We do not recommend to use Microsoft Access with smart indexing, as the ODBC configuration file requires a static drive mapping to the data source instead of a UNC path.

13.2.2.15.2 In case you want to use Microsoft Access, ensure that the drive mapping stored in the DSN configuration file is available on every machine. Configure Smart Indexing

The prerequisites for this task are:

- A form is selected that has at least two form fields.
- The database exists that contains a table with corresponding columns.
- There is an OLE DB provider for the database system you use.

To configure smart indexing, complete the following steps:

1. Select the form field that will be used as a smart index field.
2. In the vertical toolbar, at the right edge of the window, click the **Define Smart Indexing** button. The *Smart Indexing Definition* dialog displays. It shows the available fields. The smart index field is highlighted in yellow and marked with a key icon.
3. Click **Select Data Source**. The *Data Link Properties* dialog displays.
4. In the **Provider** tab, select the appropriate OLE DB provider for the type of data you want to access.
5. Click **Next**.
6. Use the **Connection** tab to specify how to connect to your data. The **Connection** tab is provider-specific and displays only the connection properties required by the OLE DB provider of your choice. For more information, click **Help**.
7. Click **OK**. The *Smart Indexing Definition* dialog redisplay.
8. Select a database table from the **Select database table** list box.
9. Click a cell in the **Database Field** column. A list containing the names of all columns of the selected table displays. Map columns to:
 - The smart index field.
 - All lookup fields. Smart indexing fills each field that is assigned to a database column.
10. Optionally, double-click each additional field that you want to use as an index field. The corresponding fields are marked with key icons. Note that search criteria are evaluated using *AND* joins, i.e. there must be a matching row for both index values.
11. A database lookup using one index term may return multiple results. In interactive mode, a dialog opens where the user can select the appropriate record. To limit the number of results, type an entry into the **Maximum # of SmartIndex Hits...** text field, or accept the default.
12. Under **Further Settings**, specify where and when smart indexing should be performed:

Use in Automatic Extraction

Smart indexing is performed during the automatic extraction step by Runtime Server.

Use in Manual Field Validation	Smart indexing is available during the extraction verification or the manual indexing step in Verifier.
Start Smart Indexing	If this option is checked, the lookup is performed on validation. For example, the lookup is performed on validation when the user presses the [Enter] key in the index field. Otherwise, the user needs to press [Alt] + [F12] to trigger smart indexing.
Reconfirm Smart Indexing	If this option is checked, the user has the opportunity to check and correct the values retrieved from the database (recommended). Otherwise, the retrieved values are validated automatically.
Accept Multiple Hits	If this option is checked, a valid lookup result may consist of multiple values.
Accept Different	If this option is checked, a valid lookup result with multiple values may have been obtained using different index fields.

13.2.2.15.3 Delete Smart Indexing Definition

When editing the form, the *Smart Index* field displays in yellow with the word **SmartIndex**.

To delete the smart indexing definition, complete the following steps:

1. Select the form field that is used as smart index field.
2. In the vertical toolbar at the right edge of the window, click the **Delete Smart Index** button. The smart index definition is deleted instantly.

13.2.3 About Editing Text Fields

Verifier provides features for editing text fields that can help users to insert and correct text faster. You can use automatic character entry when the auto-completion option 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 selected text.

A multi-line field enables line wrap and displays a vertical scroll bar, if required. To add a new line to a multi-line field, press [Ctrl] + [Enter] key.

13.3 Verification Testing

The *Verifier Test Mode* provides a complete environment for the simulation of classification and extraction verification. This includes the corresponding state evaluations and transitions.

Note: The changes of the workdoc are not saved persistently if the **Save Execution Results** to Workdoc setting is not activated.

13.3.1 About Verifier Test Mode User Interface

The user interface that displays depends on your current settings. It can be determined dynamically, using the document class and state. This leads to one of the following results:

- The classification user interface displays.
- The extraction/indexing user interface displays using the form corresponding to the current document's state.
- If no matching form is available, an empty user interface displays.
- If the user clicks the **Force Document to be validated with selected form** button, the selected form displays and is used for extraction, independent from the document properties.

13.3.1.1 Verifier Test Mode User Interface

The Verifier Test Mode user interface provides the following panes:

Forms Area	The left pane contains a set of controls for correction and manual indexing, such as form fields, labels, and viewers. Each control is assigned to a data extraction field. The viewers show the document areas that were extracted to fill the fields.
Document Area	The right pane displays the current document.
Current Input Area	This area shows a magnification of the current form field and the associated viewer. It can be used for particularly convenient viewing and editing. Note that the position of the form areas may vary according to the selected layout, or only a subset of form areas may be present.
User Info Area	This area displays the class name of the current document, the name of the current field and, if applicable, a message indicating why the current field is invalid.

13.3.1.2 Verifier Test Mode Color Coding

In *Verifier Test Mode*, color coding is used to support the quality assurance process. The appearance of the field area tells us something about the validity of the extracted data:

- A green background indicates a valid extraction result.
- A red background indicates an invalid extraction result.
- A question mark indicates illegible characters.
- A red character indicates rejects, i.e. WebCenter Forms Recognition is in doubt whether the character was read correctly.

Similar information can be obtained from the document area:

- A green highlight indicates a valid extraction result.
- A red highlight indicates an invalid extraction result.




- A yellow highlight indicates a candidate.

Depending on the current selection in the field area, document highlighting can be applied to the following items:

- Document areas associated with text fields.
- Document areas associated with entire tables.
- Document areas associated with table columns.
- Document areas associated with table rows.
- Document areas associated with table cells.





13.3.1.3 Verifier Test Mode Icons

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

Icon	Description
	Indicates the currently selected field.
	Indicates a smart index field that can be filled by a database lookup. This icon is visible only when a smart index field is selected.
	Indicates a smart index field that can be used to start a lookup. This icon is visible no matter whether a smart index field is selected or not.

13.3.1.4 Verifier Test Mode Toolbar

Use the following buttons to control form and document display during the verification procedure:

Button	Description
	If this button appears pressed, the program always keeps the focus on the document area that is associated with the currently selected field. If required, the document section displayed in the viewer changes accordingly on field change.
	If this button appears pressed, the next documents display with the currently selected magnification. Otherwise, the default magnification is used after document change.
	If this button appears pressed, the next document displays when you validate the current one. Otherwise no document change will take place, unless triggered through the navigation bar. Note that, in contrast to WebCenter Forms Recognition Verifier, valid documents are not skipped in <i>Verifier Test Mode</i> .
	If this button appears pressed, the next document is verified using the current form. Otherwise, document state and class assignment are used to determine the appropriate form on document change.

13.3.2 Modify the Settings for Force Validation

Verifier users can validate a field even if its content is invalid by definition. To modify the Verifier Test Mode settings for **Force Validation**, complete the following steps:

1. On the **Options** menu, click **Settings**.
2. On the **Verifier Mode** tab, under **Force Validation**, select one of the following settings and then click **OK**.
 - Select Permitted to enable users to validate a field even if its content is invalid by definition by pressing Enter three times.

- Select Forbidden to disable the force validation feature.

13.3.3 Modify the States for Designer Test Mode Document Routing

If you have script code associated with the *RouteDocument* event, you can use the states defined in this area to test your script in the Verifier Test Mode. This workflow is used in Verifier only. To modify the states, complete the following steps:

1. On the **Options** menu, click **Settings**.
2. On the **Verifier Mode** tab, under **VerifierTest Mode Document Routing**, modify the states and then click **OK**.

13.3.4 About Batch Renaming in Verifier

This feature enables Verifier users to rename a batch by right-clicking on it in the Verifier batch list screen. The feature provides more flexibility for Verifier users when working with exception batches and for developers when manipulating batches in custom scripts.

To ensure traceability and consistency in combination with the use of statistical tools, the following information is available for the *MoveDocument* event in the event log:

- Old batch ID
- New batch ID
- Reason for moving the document
- Document / Workflow state

13.3.4.1 Enable the Allow Batch Renaming Feature

To enable the Allow batch renaming feature, complete the following steps:

1. On the **Options** menu, click **Settings**.
2. On the **Verifier Mode** tab, select **Allow batch renaming** and then click **OK**.

13.3.5 Use the Highlighting Options

The Verifier Test Mode toolbar provides highlighting options that help you to test the verification procedure. To highlight fields and candidates, complete any of the following steps:

- To highlight all areas on the current document that were used to fill text fields, click the **Highlight all fields** button.
If the result is valid, the area highlights in green. If the result is invalid, the area highlights in red.
- To highlight only the area on the current document that was used to fill the current text field, click the **Highlight selected field** button.
If the extraction result is valid, the area highlights in green. If the extraction result is invalid, the area is highlights in red.
- To highlight the area on the current document that was used to fill the current text field, click the **Highlight all candidates** button.
If the extraction result is valid, the area highlights in green. If the extraction result is invalid, the area highlights in red. In addition, all other candidates for the current field highlight in yellow.

- To highlight all found words, click **View ▾ Highlight Words**.
- To highlight all found blocks, click **View ▾ Highlight Blocks**.
- To highlight all words whose *Visible* property is *True*, click **View ▾ Highlight Checked Words**.

Note: This option is only relevant in conjunction with table extraction or scripting using the workdoc API.

13.3.6 Control the Document Display

To control the way the form and the document display during the verification procedure, complete any of the following steps:

- To keep the focus on the document area associated with the currently selected field, click the **Keep focus on field** button.

Note: If required, the document section displayed in the viewer changes accordingly on field change.

- To display the next document with the currently selected magnification, click the **Keep zoom on switching to next document** button.

Note: Otherwise the next document displays using the default magnification.

- To move to the next document after you validated the current one, click the **Move to next document on validate** button.

Note: Otherwise the next document does not display until you click the Move to next document button. In contrast to Verifier, valid documents are not skipped in Verifier Test Mode.

- To verify the next document using the current form, click the **Force document to be validated with selected form** button.

Note: Otherwise, document state and class assignment are used to determine the appropriate form for the next document.

13.3.7 Testing Verification Form Layout

To test whether your verification form layout works with documents that have good extraction results, prepare the following document input:

- Some classified documents.

Note: To work with documents that do not cause particular problems, take the documents from the learn set.

- Create separate batches for each document class to test single forms, and another batch with mixed content to test whether the correct form displays for each class.

To test the verification form layout, complete the following steps:

Prerequisite: This procedure requires the **Allow firing of VerifierFormLoad event when in Verifier Test/Train Mode** setting activated. Do not enable this setting if the script dynamically modifies the form. Refer to [Section 4.9.5: Compatibility Settings](#).

1. Ensure that the Verifier Test Mode toolbar settings are in default state.
2. Load the documents.
3. Switch to *Verifier Test Mode*.
Because the current document is classified, the extraction /indexing user interface displays.
4. Optional. To test a single form if your document input is heterogeneous concerning the classes, on the toolbar, click the **Force document to be validated with selected form** button.
5. Optional. To resolve special problems which only occur with some documents, ensure that the **Move to next document on validate** button is not pressed.
6. Press the [Tab] key to activate the first field.
 - Indicates the currently selected field.
 - Indicates a smart index field that can be filled by a database lookup. The icon is only visible when a smart index field is selected.
 - Indicates a smart index field that can be used to start a lookup. The icon is visible no matter whether a smart index field is selected or not.
7. Check the following.
 - The field has the required size.
 - Multiline fields are available where required.
 - The associated viewer shows the appropriate document area.
 - The relevant information in the viewer is readable.
8. Optional. To modify the field content, complete any of the following sub steps:
 - To insert a new line into a multi-line field, press [Ctrl] + [Enter] key.
 - To delete all character until the end of the line, press [Ctrl] + [Del] key.
 - To delete the entire field content, press [Ctrl] + [Backspace] key.
 - To validate the current rejected character or field and move to the next invalid character or field, press [Enter] key. Depending on your current settings, this may involve a document change. Validation may also trigger smart indexing.
9. To move to the next field, press [Tab] key and then repeat the previous steps.
10. When you checked the last field, press [Enter] key to validate the document.
After the document is validated, the next document displays.
11. Repeat the previous steps for several documents to ensure that you cover all relevant cases.
12. Review the form layout in Verifier and Web Verifier.

13.3.8 Test Classes and their Assigned Forms

To test whether the classes and their assigned forms are available as expected, complete the following steps:

Prerequisite: Ensure that the Verifier Test Mode toolbar settings are in default state.

1. Prepare some unclassified documents.

Note: The quality of the documents is unimportant.

2. Load the documents.
3. Switch to *Verifier Test Mode*.

Because the current document is unclassified, the classification user interface displays.

4. To classify the document manually, in the lower pane, from the drop-down list, select a class and then press [Enter] key.

Note: Check if the classes for which the property *Visible in correction* is cleared do not appear in the list. Refer to [Section 7.9.2: Modify the Class-Level Settings](#).

The program checks whether a form is available for the selected class. If so, the extraction performs, and the document and its extraction results display using this form.

13.3.9 Testing Validation Rules

You should test whether your validation rules are functional. The list below provides an overview of possible limitations concerning field content:

- Handling of OCR rejects.
- Permitted number of characters.
- Range of valid characters.
- Restriction to uppercase / lowercase characters.
- Handling of forced validation.
- Custom validation rules implemented as WinWrap Basic scripts.

Prepare as document input:

- Classified documents.
- To run tests where you deliberately enter invalid input, you can use documents that extract correctly. You can, for instance, reuse the batches created previously to test the form layout.
- To test the handling of OCR rejects and of empty fields, create a corresponding set with OCR problems, for instance by scanning with a lower resolution.

Use the following toolbar settings in *Verifier Test Mode*:

- To test as many violations as possible with a limited number of documents, do not navigate to the next document on validation.

To run a test:

1. Load the documents and switch to *Verifier Test Mode*. Since the current document is classified, the extraction / indexing user interface displays.
2. Press the [Tab] key to activate the first field. The arrow icon appears next to the field.
3. Test the field validation rules as follows, and observe the messages that display in the status bar:
 - If the field contains OCR rejects, correct them and use the [Enter] key to confirm each corrected character. You should not be able to validate the field if rejects are forbidden.
 - If applicable, enter strings that are too long or too short. You should not be able to validate the field unless the number of characters is within the valid range.
 - If applicable, enter some invalid characters. They should be discarded automatically.
 - If applicable, enter some characters in the wrong case. They should be converted automatically.
 - If possible, enter some input that is invalid and try a forced validation by pressing the [Enter] key three times. If this is permitted, a message box displays. Select **No** to interrupt the forced validation.
4. Without validating the field, use the [Tab] key to move to the next fields and test their validation rules as well.

13.3.10 Testing Smart Indexing

You should manually test whether the Smart Indexing feature is working as expected. To do this, turn off automatic smart indexing while you are testing. You should have access to the database used for the lookup.

Prepare as document input:

- Extracted documents that have been generated with the automatic lookup turned off. To work with documents that do not cause particular problems, you can for instance take the documents from the learnset. Test each class concerned separately.

To run a test, use the default toolbar settings in *Verifier Test Mode* and complete the following steps:

1. Load the documents and switch to *Verifier Test Mode*. Because the current document is classified with an invalid extraction result, the extraction / indexing user interface displays.
2. Activate a smart index field that can be used to start a lookup. The arrow icon, the key icon, and the database icon appear next to the field. All other fields that can be filled using database lookup are marked with a database icon.
3. Run a search with the value that has been extracted, or enter a value that should return a single result. Then press [Alt] + [F12] to trigger the lookup. This should fill the lookup fields.
4. Run a search that should return multiple results. This can, for instance, be a wildcard search, using the * character to represent multiple characters, or the

? character to represent a single character. Two different results are acceptable in this case:

- If multiple results are forbidden, a dialog should display from which you can select the appropriate record. The lookup fields are then filled accordingly.
 - If multiple results are allowed, the lookup fields will be filled with multiple values. In this case, you need multiline fields to display the results properly.
5. Run a search that yields no results. You should see a corresponding error message.

If you need automatic smart indexing, do not forget to change the settings again when your tests are finished.

13.3.11 Append or Replace Text in a Field

You can correct the content of fields and table cells by appending or replacing strings selected from the document. The append feature takes the current word left of the candidate and appends the field text. It places the text in the best location, either right or left of the word.

To append or replace text in a field, complete the following steps. You can insert only one candidate per field per document verification session.

A candidate is a string that matches the learned words for that field.

1. On the toolbar, select the **Highlight all candidates** button.
2. In the left pane, select the required field.
3. On the document, click the word next to the text that you want to append.

A box appears around the word.

Note: If the word is a candidate, it highlights in green, otherwise it highlights in orange.

4. Check if the selected word fits the format analysis rules defined for the field.

Note: If a word that you expect as a candidate highlights in orange, check your format analysis rules.

5. Complete one of the following actions:

The word is a candidate

- To append the selected word in the field, right-click the box and then click **Append Field Text by Word**.
 - To replace the field content with the selected word, right-click the box and then click **Replace Field Text by Candidate**.
-

The word is not a candidate.

- To append the selected word in the field, right-click the box and then click Append Field Text by Word.
 - To replace the field content with the selected word, right-click the box and then click **Replace Field Text by Word**.
-

13.3.12 Insert Blocks of Text

Inserting large blocks of text is helpful when you have multiple word data verification elements, for fields such as address information or for cell descriptions.

To use this feature in Verifier, you must activate the **Allow immediate copying of selected area to a field or table cell** option in the settings.

For more information, refer to *Oracle WebCenter Forms Recognition Verifier Scripting User's Guide*.

To insert large blocks of text, complete the following steps:

7. Click and drag over the desired text in the image viewer.
19. Release the mouse button. A rectangle appears around the text.
20. Adjust the rectangle, if necessary, by selecting the nodes at any corner.
21. 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.

13.3.13 Testing Table Analysis and Correction

The preparations for table analysis test are essentially the same as for general extraction tests. The list below provides an overview of possible limitations concerning table content:

- Default columns.
- Cells with required entries.
- Valid column formats.
- Handling of forced validation.
- Custom validation rules implemented as WinWrap Basic scripts.

Prepare as document input:

- Classified or extracted documents.
- Create separate batches for all document classes that use common table settings.

To run a test, complete the following steps.

1. Load the documents and switch to *Verifier Test Mode*. Since the current document is classified, the extraction / indexing user interface displays.
2. The most basic test checks the presence of the table field. If some table fields are missing, you will have to improve the table analysis settings, or provide a mechanism how these cases are handled. A simple method to make sure that table fields exist (which can then be corrected) is the use of default columns.
3. The table field provides additional highlighting options:

- Clicking the button in the upper-left corner of the field highlights the entire document table.
 - Clicking a column label in the field area highlights the document column.
 - Clicking a row label in the field area highlights the document row.
 - Clicking a cell in the field area highlights the document cell.
 - Valid areas appear in green, invalid areas in red.
4. Use the highlighting options for the following tests:
- Check whether table tops and bottoms have been detected correctly.
 - Check for default columns. They must be available in valid tables. If a default column is missing, the corresponding table must be invalid.
 - Check for mandatory cells. They must be filled in valid tables. If a mandatory cell is empty, the corresponding table, column, and row must be invalid.
 - Check whether column formats have been applied correctly. If cells have an incorrect format, the corresponding table, column, and row must be invalid.
 - Check whether forced validation of invalid tables is possible.
5. The column labels' shortcut menu allows you to correct the column structure of tables. It contains the following commands:

Command	Description
Unmap Column	Removes the current column.
Swap Column	Exchanges the position of the current column and the one selected from the drop-down menu.

6. The row labels' shortcut menu allows you to correct the row structure of tables. It contains the following commands:

Command	Description
Insert Row	Inserts an empty row above the current one.
Delete Row	Deletes the current row.
Duplicate Row	Duplicates the current row.
Append Row	Appends an empty row at the bottom of the table.

7. Fill the cells with correct values and validate the table by pressing the [Enter] key.

14 Regular Expressions

14.1 What Are Regular Expressions?

Regular expressions are a generalized pattern language that WebCenter Forms Recognition uses for format analysis. With regular expressions, strings that match a specified pattern can be identified in a document's text.

Use regular expressions if you have problems getting the correct candidates for field extraction with simple expressions. Regular expressions precisely describe the candidates you are looking for.

14.2 Literal Characters in Regular Expressions

Two types of characters are used in regular expressions:

- Literal.
- Special.

Literal characters are made up of normal text characters, such as letters or numbers, and of all symbols that are not used as operators.

Special characters are symbols that can be used as operators, for example:

 * - [] ^ { } \ .

If a special character appears in a regular expression as a literal character, it must usually be marked so that it is treated as a literal character and not as an operator. To use a special character as a literal character, place a backslash in front of the symbol, or enclose the symbol with square brackets, for example:

- *
- [*]

14.3 Operators in Regular Expressions

Operators consist of one or more special characters:

Operator	Description
	Match-self Operator This operator is represented by any literal character. Its definition is included for completeness only. It matches the character itself.
	Concatenation Operator This operator is not represented by any character. It concatenates two regular expressions.
.	Match-Any-Character Operator This operator is represented by a period. It matches any single printing or non-printing character. Example: a.b matches any three-character string beginning with a and ending with b.
*	Match-Zero-or-More Operator This operator is represented by an asterisk. It is always used following a literal character or a special character. It repeats the smallest possible

Operator	Description
	<p>preceding regular expression as many times as necessary to match the pattern, including one repetition and zero repetitions.</p> <p>Examples: <code>o*</code> matches any string made up of zero or more o's. <code>fo*</code> matches <code>f</code>, <code>fo</code>, <code>foo</code> and so on.</p>
<code>\{count\}</code> <code>\{min,\}</code> <code>\{min, max\}</code>	<p>Interval Operators</p> <p>The <i>Open Interval Operator</i> is represented by <code>\{</code>. The <i>Close Interval Operator</i> is represented by <code>\}</code>. Interval operators repeat the smallest possible preceding regular expression a specified number of times: <code>\{count\}</code> matches exactly <i><count></i> occurrences of the preceding regular expression. <code>\{min,\}</code> matches at least <i><min></i> occurrences of the preceding regular expression. <code>\{min,max\}</code> matches between <i><min></i> and <i><max></i> occurrences of the preceding regular expression. The interval is invalid if <i><min></i> is greater than <i><max></i>, or any of <i><count></i>, <i><min></i> or <i><max></i> is outside the range 0 - 256.</p>
<code>[]</code> <code>[^]</code>	<p>Character Class Operators</p> <p><code>[</code> and <code>]</code> are used in regular expressions as special characters to indicate a character class. A character class matches a single character, regardless of how many characters or character ranges are defined in the character class. Characters contained in the class can be listed individually. Alternatively, character ranges can be specified using the <code>-</code> range operator. A character class can contain as many individual characters as needed, and can contain multiple character ranges.</p> <p>A negated character class can also be specified. A negated character class consists of any character except the characters and/or ranges listed in the character class. A negated character class includes the <code>^</code> metacharacter as the first character in list.</p> <p>Most special characters lose any special meaning inside a character class. The following characters still have a special meaning at certain positions within an expression: <code>]</code> closes the character class unless it is the first character. <code>\</code> quotes the next character unless it is the last character before the closing bracket. <code>-</code> represents the range operator unless it is the first character after the opening bracket or the last character before the closing bracket. Empty character classes are invalid.</p>
<code>-</code>	<p>Range Operator</p> <p>The range operator is represented by a hyphen <code>-</code>. It specifies a range of characters when used inside a character class.</p> <p>Examples: <code>a-f</code> matches any character between <code>a</code> and <code>f</code>. <code>A-Z</code> matches any capital letter. <code>a-z</code> matches any lower-case letter. <code>0-9</code> matches any number.</p>
<code>\</code>	<p>Escape Operator</p> <p>The escape operator is represented by a backslash. It is used to quote certain special characters. If you need to use the literal value of a special character within an expression, then a backslash should be inserted before that special character to signal that the following character is to be treated literally.</p>

15 Templates

15.1 About Reusing Project Settings with Templates

In data extraction, after your analysis and evaluation settings for a field are working, you can to reuse them elsewhere within or outside your project. For example, if the format analysis has been configured to find a date, it could be useful to take the same definition at any other place where dates are extracted. Therefore you can save analysis and evaluation settings as templates.

15.2 Create an Analysis or Evaluation Template for a Field

To create a template, complete the following steps:

1. Switch to *Definition Mode*.
2. In the left pane, on the **Fields** tab, select a field.
3. Complete one of the following sub steps:
 - To create an analysis template, in the right pane, on the Analysis tab, from the **Available Analysis Engines** list, select an engine.
 - To create an evaluation template, in the right pane, on the Evaluation tab, from the **Available Evaluation Engines** list, select an engine.
4. Modify the settings on the tabs below.
5. Click **Save as Template**.
6. In the **New Template** dialog box, type a name for the template and then click OK.

Note: The template is saved internally within the project.

15.3 Use an Existing Template

To use an existing template, complete the following steps:

1. Switch to *Definition Mode*.
2. In the left pane, on the **Fields** tab, select a field.
3. Complete one of the following sub steps:
 - To reuse an analysis template, in the right pane, on the **Analysis** tab, from the **Available Templates list**, select a template.
 - To reuse an evaluation template, in the right pane, on the **Evaluation** tab, from **the Available Templates** list, select a template.

Note: The tabs for analysis and evaluation settings will be hidden.

4. Optional. To adjust the template settings, complete the following sub steps:
 - a. To update the tabs for analysis or evaluation settings below with the settings from the template, click Copy Template.

- b. Modify the settings on the tabs.

Note: The modifications do not save to the selected template. To save the settings as a new template, complete the steps in Create an analysis or evaluation template for a field.

15.4 Display the Available Templates

Templates are stored and managed on the project level. To display the templates, complete the following steps:

1. Switch to *Definition Mode*.
2. In the left pane, on the **Classes** tab, right-click the project and then click either **Show Analysis Templates** or **Show Evaluation Templates**.

15.5 Modify a Template

To modify a template, complete the following steps:

1. Switch to *Definition Mode*.
2. In the left pane, on the **Classes** tab, right-click the project and then click either **Show Analysis Templates** or **Show Evaluation Templates**.
3. Under **Available Templates**, select a template.
4. Modify the settings in the tab.
5. On the **File** menu, click **Save Project**.

Note: All fields using the template immediately work with the new settings.

15.6 Rename a Template

You can only rename templates that are currently not in use. To rename a template, complete the following steps:

1. Switch to *Definition Mode*.
2. In the left pane, on the **Classes** tab, right-click the project and then click either **Show Analysis Templates** or **Show Evaluation Templates**.
3. To rename a template, complete the following sub steps:
 - a. Under **Available Templates**, select a template and then click **Rename**.
 - b. In the **Rename Template** dialog box, type the new name and then click **OK**.

15.7 Delete a Template

You can only delete templates that are currently not in use. To delete a template, complete the following steps:

1. Switch to *Definition Mode*.

2. In the left pane, on the **Classes** tab, right-click the project and then click either **Show Analysis Templates** or **Show Evaluation Templates**.
3. Under **Available Templates**, select a template and then click **Delete**.
4. In the confirmation dialog box, click **Yes**.

15.8 Export Templates to another Project

You can export analysis templates as files to use them in other projects. This feature is not supported for evaluation templates. To export templates, complete the following steps:

1. Switch to *Definition Mode*.
2. In the left pane, on the **Classes** tab, right-click the project and then click either **Show Analysis Templates** or **Show Evaluation Templates**.
3. In the right pane, on the **Analysis Templates** tab, click **Export**.
4. In the Export Templates dialog box, clear the check boxes for the templates you do not want to export and then click **OK**.

Note: All selected templates are exported into one file with **.exp** extension.

5. In the Save As dialog box, in the **File name** field, type a name and then click **Save**.

15.9 Import Templates from another Project

To import templates, complete the following steps:

1. Switch to *Definition Mode*.
2. In the left pane, on the **Classes** tab, right-click the project and then click either **Show Analysis Templates** or **Show Evaluation Templates**.
3. In the right pane, on the **Analysis Templates** tab, click **Import**.
4. In the Open dialog box, select a file with **.exp** extension and then click **Open**.

Note: The program automatically recognizes the template type contained in the export file. If the type is wrong, the template is not imported.

16 Document Export

16.1 About setting up the document export

Without a connector or an export script, the documents are written to the export directory specified in the settings after processing.

The following export connectors are available:

- ITA-DMS connector
- SAP FI module connector
- Lotus Notes connector
- Microsoft Exchange connector

You can use these export methods together or independently. If you need connectors, contact your support representative.

Individual document export options need to be implemented using scripts. For more information, refer to *Create a Script for Document Export* in the *Oracle WebCenter Forms Recognition Scripting User's Guide*.

17 Regional and Currency Settings

The following table lists settings for region and currencies used in *Validation* and *Output Formatting*:

Region	Currency	Supported Symbols
Dutch	Euro, Guilder	NLG, fl, €
Dutch-Belgium	Euro, Franc	BEF, BF, €
English	Euro, Dollar	USD, \$, €
English-Australia	Euro, Dollar	AUD, \$, €
English-Canada	Euro, Dollar	CAD, \$, €
English-New Zealand	Euro, Dollar	NZD, \$, €
English-United Kingdom	Euro, Pound	GBP, €
English-United States	Dollar	USD, \$, €
French	Euro, Franc	FRF, F, €
French-Belgium	Euro, Franc	BEF, BF, €
French-Canada	Euro, Dollar	CAD, \$, €
French-Switzerland	Euro Franc	CHF, SFR, €
German	Deutsche Mark, Euro	DEM, DM, €
German-Liechtenstein	Deutsche Mark, Euro	DEM, DM, €
German-Luxembourg	Deutsche Mark, Euro	DEM, DM, €
German-Swiss	Euro, Franc	CHF, SFr, €
Italian	Euro Lira	ITL, L, €
Italian-Swiss	Euro, Franc	CHF, SFr, €
Neutral	Euro	€
Norwegian	Euro, Kroner	NOK, k
Polish	Euro Zloty	€
Russian	Euro	€
Spanish	Peseta	ESP, pta
Spanish-Mexico	Peso	MXN, ls
Spanish-Modern	Peseta	\$, ESP, pta
Swedish	Euro, Krona	SEK, KR, €