Oracle® Fusion Middleware

WebCenter Forms Recognition Designer User's Guide

14c (14.1.1.0.0)

F73576-01

August 2023

Describes how to create, define, and modify projects

Oracle Fusion Middleware Oracle WebCenter Forms Recognition Designer User’s Guide, 14c (14.1.1.0.0)

F73576-01

Copyright © 2009, 2023, Oracle and/or its affiliates.

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, software documentation, data (as defined in the Federal Acquisition Regulation), 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 embedded, installed, or activated on delivered hardware, and modifications of such programs) and Oracle computer documentation or other Oracle data delivered to or accessed by U.S. Government end users are "commercial computer software," "commercial computer software documentation," or "limited rights data" pursuant to the applicable Federal Acquisition Regulation and agency-specific supplemental regulations. As such, the use, reproduction, duplication, release, display, disclosure, modification, preparation of derivative works, and/or adaptation of i) Oracle programs (including any operating system, integrated software, any programs embedded, installed, or activated on delivered hardware, and modifications of such programs), ii) Oracle computer documentation and/or iii) other Oracle data, is subject to the rights and limitations specified in the license contained in the applicable contract. The terms governing the U.S. Government's use of Oracle cloud services are defined by the applicable contract for such services. 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®, Java, and MySQL are registered trademarks of Oracle and/or its affiliates. Other names may be trademarks of their respective owners.

Intel and Intel Inside 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, Epyc, and the AMD 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.

Table of Contents

What is Oracle WebCenter Forms Recognition Designer? 7
Specify the Administrator Password 7
Roles...7
About Groups 8
About Users 9
About Windows User Import 14
About Modes 14
What are Workdocs?.. 32
The Document Pane 32
Switch Between the Workdoc and the Page Definition View 33
Document Processing 33
About Learning 33
The Learning Toolbar... 34
Compress a Learnset... 34
About Highlighting 34
The Highlighting Toolbar 35
Print Documents 35
What are Project Files? 35
Open a Project created with an Earlier Version 36
Create a Project.. 36
About the GUI Character Set 36
Save a Project Without Learn Data 37
About the Save Compress Feature 37
Save Compress a Project 37
Open a Project 37
Version Control 38
About the Project Structure 40
About the Structure of Non-Associative Projects 40
Project Planning 41
About Classification 45
About Advanced Classification.. 46
Prepare Sample and Test Documents 46
Prepare the Project 46

About Classes 46
About Classification Methods 49
Select Classification Methods 49
Project-Level Standard Classification 49
About Stretching 54
About Calibration 55
Classification Learnsets 55
About the Brainware Classify Engine 60
About the Layout Classify Engine 62
About the ASSA Classify Engine 62
About the Brainware Layout Classification Engine 63
About the Forms Classify Engine 66
About the Phrase Classify Engine 69
About the Language Classify Engine 75
About the Image Size Classification Engine 78
About the Template Classify Engine 78
Engine-Level Absolute Results 79
About Classification Testing 80
Classification-related columns on the Statistic View tab 82
Extraction-related columns in the lower left pane 82
Classification Problems 84
Recognition 85
OCR (Optical Character Recognition) 85
OMR (Optical Mark Recognition) 85
Barcode Recognition 85
FineReader OCR Engines 86
FineReader11 Barcode Engine 87
QualitySoft Barcode Engine 88
Cleqs Barcode Engine 89
About Data Extraction 92
About Threshold and Distance 93
Field Types 93
Create a Field 94
Create a Table Field 94
Specify a Display Name for a Field 94
Delete a Field 94

Check the Analyzed Table Content 95
Rename a Field 95
Specify the Analysis Engine 95
About Project-Level Recognition Settings 95
Modify the Field-Level Recognition Settings 96
Specify the OCR Settings 97
About Despeckling 97
Zone Analysis Engine 125

About Validation 131

About the Standard Validation Engine 131
About Validation Settings 131
About the Force Validation Feature 131
About Derived Validation 132
About Validation Rules 132
About document Validation 132

About Field Evaluation 147

Configure the Brainware Extraction Engine 147
Brainware Field Extraction Engine for Generic Field Extraction 148
About Training Header Fields in Normal Train Mode 150
Train Header Fields in Normal Train Mode 150
About Learning the Extraction 151
About Creating Learnsets 151
About Learning 153
About Testing the Extraction 154
Extraction Optimization 155
Text Field Candidate Evaluation 157
Table Field Candidate Evaluation 158

About Supervised Learning 158

Considerations and Constraints for Supervised Learning Workflow 159
Apply the Supervised Learning Workflow 159
Example for a Supervised Learning Workflow 159
Train the Base Class 163
Establish Multiple Classification Results Using a Script 164
Training of Other Fields 164
About Creating Derived Document Classes 164

Create Derived Document Classes with the Supervised Learning Workflow 165
About the Supervised Learning Workflow Process 165
Add Supervised Learning to Brainware Layout Classification 165
Specify the Default Classification Result 166
ASE Pool Configuration 166
Options for Creating Derived Classes in Verifier 167
About Quality Assurance with Verifier.. 167
About Verification Projects... 168

Add a Viewer 176

Verification Testing 181

What are Regular Expressions? 191

Literal Characters in Regular Expressions 192
Operators in Regular Expressions.. 192

About Reusing Project Settings with Templates 197

Create an Analysis or Evaluation Template for a Field 197
Use an Existing Template... 197
Display the Available Templates 197
Modify a Template... 198
Rename a Template... 198
Delete a Template 198
Export Templates to Another Project 198
Import Templates from Another Project... 199
About Script Encryption 199
Activate Script Encryption 199
Debug an Encrypted Script.. 200
Decrypt a Project... 200
Use Breakpoints in Encrypted Scripts 201
Bypass Debugging Issues... 201
Define a Complex Sequence for Class script Dependencies 201

Regional and Currency Settings 202

What is Oracle WebCenter Forms Recognition Designer?

Oracle WebCenter Forms Recognition Designer is a component of the WebCenter Forms Recognition product suite.

Within Designer, you create, define, and modify your projects to customize the automatic data extraction process for paper and electronic documents. You can also test system settings prior to production.

You define how to classify documents relevant to 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 pass to the Runtime Server for production processing.

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

Specify the Administrator Password

As an administrator, your first task is to specify a password for each existing project. To change your password, complete the following steps.

1. Load an existing project using **Administrator** as user name.
2. From the **Options** menu, select **Users, Groups and Roles**.
3. On the **Project Authentication Properties** dialog box, on the **Change Password** tab, enter the password information in the appropriate fields.

   Note: The minimum character length for a password is 7, the maximum character length is 20. The password must include both alphanumeric and numeric characters.

4. Click **Apply** and then click **OK**.

Roles

WebCenter Forms Recognition provides the following predefined roles.

Administrator (ADM)

This role encompasses the permissions of all roles. In WebCenter Forms Recognition, administrators design and maintain projects. They also create user accounts, passwords, groups, and assign users to groups. Also, the administrators assign one or more roles to each user group.

Note: The user role is always more prominent than the group role.

Authorization for External Batches (AEB)

Users with this role can submit and handle documents using OracleWFR API.

Supervised Learning Manager (SLM)

Users with this role can define, modify, and maintain the learnset. This functionality is available in Verifier only.

Supervised Learning Verifier (SLV)

Users with this role can collect and manage local training data, and can propose learnset candidates to improve the performance. This functionality is available in both Verifier and

Web Verifier.

Verifier (VER)

Users with this role can verify the documents that the Runtime Server could not automatically process. This functionality is available in both Verifier and Web Verifier.

Verifier Filtering (FLT)

Users with this role can access and change the batch filtering properties in both Verifier and Web Verifier.

Verifier Settings (SET)

Users with this role can access and change the Verifier and Web Verifier settings.

About Groups

Users inherit the roles of their assigned groups. A group must have one or more assigned roles.

Add a Group

By default, the first time the administrator logs in, the only existing user group is ADM, and the only member of the ADM group is the Administrator user. After you have established user groups, you are ready to assign users to them. To add a new group, complete the following steps.

1. From the **Options** menu, select **Users**, **Groups and Roles**.
2. On the **Project Authentication Properties** dialog box, on the **Groups** tab, click **New**.
3. In the **Add New Group** dialog box, in the **Name** field, enter a name for the new group.

   Note: The maximum character length for a group name is 20.

4. Select one of the following group roles, and then click **OK**.
   - Administrator
   - Supervised Learning Manager
   - Supervised Learning Verifier
   - Verifier
   - Oracle WFR API access
5. Under **Role Assignments**, add or remove the required roles, click **Apply** and then click **OK**.

Delete a Group

To delete a group, complete the following steps.

1. From the **Options** menu, select **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.**

Add a Role to a Group

When you first create a group, you can only assign one of the existing roles to it. However, you can assign or remove roles from an existing group. To add a role to a group, complete the following steps.

1. From the **Options** menu, select **Users**, **Groups and Roles**.
2. On the **Project Authentication Properties** dialog box, on the **Groups** tab, under **Available Roles**, select a role and then click **Add**.
3. Click **Apply** and then click **OK**.

Remove a Role from a Group

To remove a role, complete the following steps.

1. From the **Options** menu, select **Users**, **Groups and Roles**.
2. On the **Project Authentication Properties** dialog box, on the **Groups** tab, under **Assigned Roles**, select a role and then click **Remove**.
3. Click **Apply** and then click **OK**.

About Users

To open a project in Designer, the user must log on with a user ID and a password. You

can create the following types of users.

- Project users that are only available in the project.
- User created and updated by script.
  For more information, see About users imported by script.
- Users imported from Active Directory.

  For more information, see About Windows user import.
- Windows authenticated users, includes project users and script imported users. For more information, see About Windows authentication.

Notes

- All Designer users must have local system admin rights.
- A user must be a member of one or more groups.

User Account Settings

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.

- Allow Windows Authentication
- Allow Database Authentication

Create a User

To create a user, complete the following steps.

Note: The **Enable password remembering for user logon dialog** option is deprecated and no longer taken into account.

1. From the **Options** menu, select **Users, Groups and Roles**.
2. In the **Project Authentication Properties** dialog box, on the **Users** tab, click **New**.
3. In the **Add New User** dialog box, complete the following substeps and then click **OK**.
    1. In the **Name** field, enter the user name.

       Note: The maximum character length for a user name is 17.

    2. In the **Password** field, enter the password.

       Note: The minimum character length for a password is 7, the maximum character length is 20. The password must include both alphanumeric and numeric characters.

4. Optional. In the **Local Project** field, enter or browse to the path of a local project.
5. Optional. To specify the knowledge source location, complete one of the following substeps.
    1. In the **Common Folder** field, enter or browse to the path of a directory.
    2. 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 on 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. Select **Allow Windows Authentication**.
10. For Designer and Verifier users, select **Allow Database Authentication**.

    Note:

    You do not need to select this option for Web Verifier users, as Web Verifier always authenticates against database users. For more information, see About Database Authentication.

11. Click **Apply** and then click **OK**.

Delete a User

To delete a user, complete the following steps.

1. From the **Options** menu, select **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**.

Assign a User to a Group

To assign a user to a group, complete the following steps.

1. From the **Options** menu, select **Users**, **Groups and Roles**.
2. On the **Project Authentication Properties** dialog box, on the **Users** tab, select a user.
3. From the **Available Groups** list, select a group and then click **Add**.
4. Click **Apply** and then click **OK**.

Remove a User from a Group

To remove a user from a group, complete the following steps.

1. From the **Options** menu, select **Users**, **Groups and Roles**.
2. On the **Project Authentication Properties** dialog box, on the **Users** tab, select a user.
3. From the **Assigned Groups** list, select a group and then click **Remove**.
4. Click **Apply** and then click **OK**.

Change a Password

You can change password only for users created within the project file. Users created by the security update feature do not display in the user name list. To change password, complete the following steps.

1. From the **Options** menu, select **Users, Groups and Roles**.
2. In the **Project Authentication Properties** dialog box, on the **Change Password** tab, from the **User name** list, select the required user.
3. Optional. To display the passwords, select **Decode password**.
4. In the **Old password** field, enter the current password.
5. In the **New password** field, enter the new password.

   Note: The minimum character length for a password is 7, the maximum character length is 20. The password must include both alphanumeric and numeric characters.

6. In the **Confirm new password** field, reenter the new password.
7. Click **Apply** and then click **OK**.
8. Optional. To activate the **Update passwords in Database** option on the **Change Password** tab in Verifier and Designer, select the **Allow Database Authentication** option.

About Database Authentication

Designer stores user credentials either in the project file or in the database.

Designer and Verifier can use the users contained in the database to authenticate to the project if the **Allow Database Authentication** option is enabled.

Designer and Verifier first attempt to authenticate against the user stored in the project file. If the user is not available in the project, and **Allow Database Authentication** is enabled, the application attempts to authenticate against database users.

Web Verifier always authenticates against database users.

About Exporting and Importing Users Between Projects

You can export all users and groups of a project to a file and then import them into another project.

Export Users and Groups

To export all users and groups, complete the following steps.

1. From the **Options** menu, select **Users, Groups and Roles**.
2. In the **Project Authentication Properties** dialog box, click **Export to File**.
3. In the **Save As** dialog box, in the **File name** field, enter the file name and then click **Save**.

   Note: Designer saves the file as an SEC file.

Import Users and Groups

To import users and groups exported from another project, complete the following steps.

Note: The import function replaces all existing users and groups with those from the selected SEC file.

1. From the **Options** menu, select **Users, Groups and Roles**.
2. In the **Project Authentication Properties** dialog box, on the **Users** tab, click **Import from File**.
3. In the **Open** dialog box, select the required SEC file and then click **Open**.
4. Click **Apply** and then click **OK**.

About Exporting Users to the Database

You can export all users and groups of a project to the WebCenter Forms Recognition database to make them available for all WebCenter Forms Recognition programs.

Export Users and Groups to the WebCenter Forms Recognition Database

To make users, and groups of the current project available in the WebCenter Forms Recognition database, complete the following steps.

Note: The export only affects database users and groups which have the same name than those defined in the project.

1. From the **Options** menu, select **Users, Groups and Roles**.
2. In the **Project Authentication Properties** dialog box, on the **Users** tab, click **Export to database**.

About Users Imported by Script

To create and update users, use the security update feature. For more information, see Import users and groups using the security update feature.

Note: Users imported by script are only stored in the database and not in the project.

Import Users and Groups Using the Security Update Feature

To import users and groups using the security update feature, complete the following steps.

1. From the **Options** menu, select **Users, Groups and Roles**.
2. On the **Users** tab, select **Allow Database authentication** and then click **OK**.

3. Switch to **Definition Mode**.
4. On the toolbar, click the **Show/hide script** button.
5. From the **Object** list, select **ScriptModule**.
6. From the **Proc** list, select **UpdateSystemSecurity**.
7. Modify the script according to your needs.

   Note: For more information, see “UpdateSystemSecurity” in the *Oracle WebCenter Forms Recognition Scripting User's Guide*.

8. Configure one RTS instance to execute **Update System Security** with this project.
9. Run the RTS instance.

About Windows Authentication

The Windows-based user authentication feature simplifies the logon procedure.

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

If the currently logged in Windows user does not exist in the WebCenter Forms Recognition database, the WebCenter Forms Recognition login dialog displays.

Note: You cannot use your Windows user credentials in the Verifier and Designer login dialog box.

Enable Windows Authentication

To enable Windows authentication for a project in Designer, complete the following steps.

1. Open the project.
2. From the **Options** menu, select **Users, Groups and Roles**.
3. In the **Project Authentication Properties** dialog box, on the **Users** tab, complete the following substeps.
    1. Select **Allow Windows Authentication**.
    2. Click **OK**.
4. From the **File** menu, select **Save Project**.

Enable Windows Authentication for Web Verifier

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

1. Complete one of the following substeps.
    - Import the required Windows user accounts.

      For more information, see Import Windows users.

    - Import the required users by script.

      For more information, see Import users and groups using the security update feature

2. Configure Windows authentication for Web Verifier in IIS.

   Note: Only Domain users can access the Web Verifier. If the currently logged in windows user cannot log on to Web Verifier, the IIS login dialog displays so that the user can log on with windows credentials.

About Windows User Import

Designer provides the ability to import Windows users from Active Directory. For more information, see Import Windows users in Designer.

Note: We recommend to import users by script as described in Import users and groups using the security update feature.

Import Windows Users

To import Windows users, complete the following steps.

1. From the **Options** menu, select **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: Designer 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.
8. For each imported Windows user, assign the required groups.
9. In the **Project Authentication Properties** dialog box, on the **Users** tab, select **Allow Database Authentication** and then click **Export to Database**.
10. Save the project.

About Modes

Designer provides several operation modes. Each mode provides particular tools to design a project.

When you switch from one mode to another, the toolbar and the panes change to provide the corresponding tools and controls.

About the Document Input Selection Mode

Use this mode for the following purposes.

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

Switch to Document Input Selection Mode

To switch to Document Input Selection mode, complete the following steps.

1. On the toolbar, click the **Switch to Document Input Selection** button.

   Note: Depending on the current input type the appearance of this button varies.

2. To select the input type, from the **Switch to Document Input Selection** list, select one of the following options.
    - To use directories as input source, select **Directory**. The button changes to 📁
    - To use batches as input source, select **Batch**. The button changes to 📂 This menu item is not available if you did not specify the **Batch-Root Directory** on the **Input Mode** tab in the settings.
    - To use the Learn Set as input source, select **Learn Set**. The button changes to 🛢. This menu item is not available if you did not specify the **Learn Set Manager Base Directory** on the **Train Mode** tab in the settings.

Specify the Learn Set Location

To specify the learnset location, complete the following steps.

1. From the **Options** menu, select **Settings**.
2. In the **Settings** dialog box, on the **Train Mode** tab, in the **Base directory** field, enter or browse to the learnset directory and then click **OK**.

Specify the Input Mode

To specify the input mode, complete the following steps.

1. From the **Options** menu, select **Settings**.
2. In the **Settings** dialog box, select **Use database as documents and statistics source**.
3. On the **Input Mode** tab, under **Import Options**, to select the file format, complete one of the following steps.
    - Select **Image & Electronic Documents** and enter the required file extension, such as TIF or PDF.
    - Select **Workdoc Files (WDC)** to use workdocs without image files.
4. Optional. To create a new job, complete the following substeps.
    1. Click the **Create Job** button.
    2. In the **Oracle WebCenter Forms Recognition Designer** dialog box, enter a job name and then click **OK**.
5. Select the required job from the **Select Job** list.
6. Optional. Select **Save Execution Results to Workdoc** if you want to permanently save the result of the processing done in Designer to the workdoc.

Specify the Location of Batches

Note: This option is no longer in use.

You must specify the batch location before you can select batches as input type. To specify the location of batches, complete the following steps.

1. From the **Options** menu, select **Settings**.
2. In the **Settings** dialog box, on the **Input Mode** tab, in the **Batch-Root Directory** field, enter or browse to the required directory.
3. In the **Image-Root Directory** field, enter or browse to the required directory. **Note:**

In general, both root directories should be the same.

Input Type Directory

Use directories as 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.

Create a Batch from Files in a Directory

To create a batch from files in a directory,complete the following steps.

1. In the top left pane, double-click a directory.
2. From the **Edit** menu, select **Import New Batch from Directory**.

   WebCenter Forms Recognition inserts the selected batch in the batch directory.
3. Optional. To display the batch, from the **Switch to Document Input Selection** list, select **Batch**.

Input Type Batch

Runtime Server creates and processes batches. For classification and extraction design, use batches as input type.

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.

Create a New Empty Batch

To create a new empty batch, complete the following step.

- From the Edit menu, click Batch > Create Batch.

Rename a Batch

To rename a batch, complete the following steps.

1. In the top left pane, right-click the batch and then click **Rename**.
2. In the **Name** field, enter a new name and then press Enter.

Change the Batch State

To change the batch state, complete the following steps.

1. In the top left pane, right-click the batch and then click **Change State**.
2. From the **State** list, select a state and then press Enter.

Change the Batch Priority

To change the batch priority, complete the following steps.

1. In the top left pane, right-click the batch and then click **Change Priority**.
2. From the **Priority** list, select a priority and then press Enter.

Close a Batch

If you open a batch, Designer locks the batch. To unlock the batch and enable further processing, for example in Runtime Server, you need to close the batch. Complete the following step.

- Right-click the batch and then click Close.

Delete a Batch

To delete a batch, complete the following step.

1. In the top left pane, right-click the batch and then click **Delete Batch**.
2. In the **Delete batch** dialog box, click **Yes**.

Add a Folder

To add a folder to a batch, complete the following steps.

1. In the top left pane, double-click the batch.
2. Right-click the batch and then click **Add Folder**.
3. In the **Add Folder** dialog box, enter a folder name and then click **OK**.

Rename a Folder

To rename a folder, complete the following steps.

1. In the bottom left pane, select the folder.
2. From the **Edit** menu, select **Folder > Rename**.
3. In the **Rename Folder** dialog box, in the **Name** field, enter a new name and then click **OK**.

Delete a Folder

To delete a folder, complete the following steps.

1. In the bottom left pane, select the folder.
2. From the **Edit** menu, select **Folder > Delete**.
3. In the **Delete Folder** dialog box, click **Yes**.

Add a Document to a Folder

To add a document to a folder, complete the following steps.

1. In the bottom left pane, select the folder.
2. From the **Edit** menu, select **Folder > Add Document**.
3. In the **Add Workdoc to Folder** dialog box, select a workdoc file and then click **Open**.

Change the Document State

To change the document state, complete the following steps.

1. In the bottom left pane, select the document.
2. From the **Edit** menu, select **Document > Change State**.
3. From the **State** list, select a state and then press Enter.

Delete a Document

To delete a document, complete the following steps.

1. In the bottom left pane, select the document.
2. From the **Edit** menu, select **Document > Delete**.
3. In the **Delete Document** dialog box, click **Yes**.

Move a Folder to Another Batch

To move a folder to another batch, complete the following steps.

1. From the bottom left pane, drag the folder to the required batch in the top left pane.
2. In the **Move Folder** dialog box, click the required button.

Move a Document to Another Folder

To move a document to another folder, complete the following step.

- In the bottom left pane, drag the document to the required folder.

Input Type Learn Set

To train classification and extraction, you need a learnset. Select Learn Set as input mode 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 and 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.

Display the Classification Learnset

The classification learnset is a set of sample documents used for classification. To display the classification learnset, complete the following steps.

1. In the top left pane, select a class.
2. Right-click the class and then click **View Classification Learn Set**.

Display the Extraction Learnset

To display the extraction learnset, complete the following steps.

1. In the top left pane, select a class.
2. Right-click the class and then click **View Extraction Learn Set**.

Remove a Document form the Learn Set

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 Learn Set**.
3. Re-learn the documents.

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.

About 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.

The Processing tab of project settings in Designer includes Runtime Server and Designer import options that allow you to control the automatic conversion of documents during import into the WebCenter Forms Recognition system. These options affect the following import processes.

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

Automatic Document Conversion Settings

Convert images into

Select this option to specify the image format to which Designer and Runtime Server convert the imported files. The default and recommended image format is TIF G4 compr, which supports multiple pages in one file.

Automatically convert resolution to

Use this option to specify the resolution for the converted files. The default and recommended value is 300 * 300.

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.

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”.

About the Definition Mode

In Definition Mode you can perform the following tasks.

- Design the project settings
- Establish classification and data extraction

- Generate a default export file
- Execute an export during classification and extraction
- Test single documents or an entire document set

Switch to Definition Mode

To switch to Definition Mode, complete the following step.

- On the toolbar, click the Switch to Definition Mode button.

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 modify the processing delay, complete the following steps.

1. From the **Options** menu, select **Settings**.
2. In the **Settings** dialog box, on the **Definition Mode** tab, under **Definition Mode**, set the time delay.

About Non-Western Languages

Configure the Phrase Classification, Table Analysis and Format Analysis engines using phrases or strings in the desired non-western language on the corresponding non-western PC or on any PC where this language is the default system’s 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 WebCenter Forms Recognition Extraction engine, the WebCenter Forms Recognition and ASSA Classification engines, the WebCenter Forms Recognition Layout Classification engine.

While the support option for non-western languages always looks for a one-to-one character transformation, the CJKT 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 `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
- Russian
- Chinese, Japanese, Korean, and Taiwanese (CJKT)

Designer applies an automatic phonetic translation from non-western languages into English and then an unambiguous conversion back to the original language. The conversion applies for the following workflow steps in WebCenter Forms Recognition.

- Classification
- Extraction
- Validation
- Learning

You can execute these processes on a non-native PC, 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 engines allow you to enter phrases and format strings 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.

Activate Multi-Byte Encoding for CJKT Languages

If you use CJKT languages in your project, you need to activate multi-byte encoding. Complete the following steps.

1. From the **Options** menu, select **Settings**.
2. In the **Settings** dialog box, on the **Definition Mode** tab, select **Activate support of non-western languages**.
3. Select **Use multi-byte encoding** and then click **OK**.
4. Save and re-learn the project.

Enable Advanced Options for Additional Language Conversion Settings

Use the `LanguageSupportWorkflowSettingsVisible` registry key to enable or disable advanced options for additional language conversion settings in Designer. 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 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, enter `LanguageSupportWorkflowSettingsVisible` and then click **OK**.
5. Right-click the **LanguageSupportWorkflowSettingsVisible** key and then click **Modify**.
6. In the **Edit DWORD (32-bit) Value** dialog box, in the **Value data** field, complete one of the following steps and then click **OK**.
    - To disable the feature, enter **0**.

  - To enable the feature, enter **1**.

Copy a Non-Western Language Phrase from a Document

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.

Definition Mode User Interface

Classes and Fields pane

The Classes tab displays the project classes. The Fields tab displays the fields that Designer extracts from the documents.

Document pane

The document pane displays the selected document.

Properties pane

The properties pane displays the properties of the selected project, class, or field.

To display the properties pane, on the toolbar, click the Show/hide class and field properties button.

Display the Project Properties

To display the project properties, complete the following steps.

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

Project Properties Pane

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.

Classification tab

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

Advanced tab

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

Supervised Learning tab

For more information, see About Supervised Learning.

Display the Class Properties

To display the class properties, complete the following steps.

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

Class Properties Pane

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.

Display the Field Properties

To display the field properties, complete the following steps.

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

Field Properties Pane

Analysis tab

The Analysis tab contains the available templates, the available analysis engines and their settings.

Evaluation tab

The Evaluation tab contains the available evaluation engines and their settings.

Field tab

The Field tab contains the OCR settings for the selected field.

Validation tab

The Validation tab contains the available validation engines and their settings.

About the 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 the document selection mode. Although learning is also available in Definition Mode, only the Train Mode allows you to create learnsets.

Switch to Normal Train Mode

To switch to Normal Train Mode, complete the following step.

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

Specify the Train Mode Settings

To specify the Train Mode settings to create a learnset for classification and extraction, complete the following steps.

1. From the **Options** menu, select **Settings**.
2. In the **Settings** dialog box, on the **Train Mode** tab, select one or all of the following options:
    - To create a learnset or add documents to it, select the **Add trained documents to the learnset** check box. To use the Train Mode for demonstration purposes, clear the **Add trained documents to the learnset** check box.
    - To classify only new document that are added to the learnset for data extraction, select **Execute classification for new documents only**. Use this option if your sample documents do not belong to the same classes. If you work with sorted input, clear the **Execute classification for new documents only** check box.
    - To use previously extracted values from the document and pre-fill form fields with them during learning in **Normal Train Mode**, select the **Review in normal train mode previously extracted values** check box.
3. In the **Base Directory** field, verify or change the path to the directory where Designer stores the learnset documents.

   Note: Each time you add a sample document to the learnset of a class, Designer copies the document to this directory.
4. Optional. To create a backup before learning new documents, select the **Always automatically create a learnset backup before applying learning** check box.

Train Mode User Interface

The Train Mode user interface enables you to train classification and extraction.

Classes and Fields pane

To add a document automatically to the classification learnset, select the Classes tab. In the Classes tab, you can perform the following tasks.

- Manually assign a document to a class
- Modify the proposed class assignments

To add a document automatically to the extraction learnset, select a field on the Fields tab. In the Fields tab, you can perform the following tasks.

- Manually select the correct field for data extraction from several candidates
- Modify the proposed candidates for data extraction

Document pane

The document pane displays the current document.

About the 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.

About 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, Designer creates a new class automatically.

Automatic Supervised Learning performs the following actions.

- Extracts the supplier information.
- Creates a new class using the supplier name if the document is added to the learnset.
- Learns the new class.
- Migrates the local project to the global learnset.

Enable the Verifier Train Mode

To enable the Verifier Train Mode, complete the following steps.

1. From the **Options** menu, select **Settings**.
2. In the **Settings** dialog box, on the **Verifier Train Mode** tab, select **Enable automatic supervised learning workflow for this project**.
3. Create a verification form.

Note: For more information, see About creating verification forms.

Specify the Verifier Train Mode Settings

To specify the Verifier Train Mode settings, complete the following steps.

1. From the **Options** menu, select **Settings**.
2. To create new classes using the supplier’s name and to create a learnset, select the **Apply local classification and extraction automatically** check box.
3. Select the **Prompt if script forces or rejects insertion to the Learnset** check box.

4. Under **Put document to local Learnset**, select one of the following options.

Situation	Steps
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.	• Select **Only if adding activated by a user**.
Learn documents only when they are initially unclassified.	• Select **Automatically if more than [x]% invalid fields** and clear **Always prompt before adding**.
Automatically verify that a document was added to the learnset if the number of documents in the class was updated and to learn documents when the specified percentage of fields is invalid before they are manually corrected.	• Select **Automatically if more than [x] % invalid fields** and **Always prompt before adding**.

5. Under **Learn new documents**, select one of the following options.

Situation	Steps
Initiate learning only if the user requests it.	• Select **Only by user request**.
Initiate learning for every batch in the project each time any batch is closed.	• Select **Before batch closing**.
Initiate learning any time a document is added to the learnset.	• Select **Immediately**.

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

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.

Train the Learn Set

To train the learnset, 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.

Verifier Train Mode User Interface

Form pane

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.

Current Input pane

The lower middle pane displays a magnification of the current form field and the associated viewer. You can use it for particularly convenient viewing and editing.

Note: 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 pane

The lower pane 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.

Document pane

The right pane displays the current document.

Verifier Train Mode Toolbar

The Verifier Train Mode toolbar provides navigation, highlighting, zooming, and validating functions.

Button	Description
←	Move to previous page
→	Move to next page
▤	Highlight all fields
▤	Highlight selected fields
▤	Highlight all candidates
↕	Maximize image height

↔	Maximize image width
✥	Maximize image size
🔍	Zoom in
🔍	Zoom 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

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 used in definition or Train Mode

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

Switch to Runtime Mode

To switch to Runtime Mode, complete the following step.

- On the toolbar, click the Switch to Runtime Mode button.

Specify the Runtime Mode 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 specify the Runtime Mode settings, complete the following steps.

1. From the **Options** menu, select **Settings**.
2. In the **Settings** dialog box, on the **Runtime Mode** tab, under **Workflow Settings**, select any of the following settings.
    - To execute classification automatically, select **Execute Classification**.
    - To execute extraction automatically, select **Execute Extraction**.

    - To redo OCR with the selected settings, select **Redo OCR on Documents**.
    - To execute the export automatically, select **Execute Export**.
3. Under **Export Options for Definition and Runtime Mode**, select any of the following options.
    - To trigger the ExportDocument script event that contains export instructions, select **Trigger script-based export**.
    - To copy processed documents to the export directory, select **Copy documents to export directory**.
    - To generate an EXP export ASCII file with the processing results in the export directory, select **Generate default export file for each batch**.
    - To generate a PDF file with the processing results in the export directory, select **Generate PDF file on export**.
4. In the **Export** directory field, verify or change the directory where Designer stores the results of the export processing.

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 document classes, each event for that document classes also executes 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 appropriate.

Change the settings for Inheritance Chronology event sequences only if scripts exist for those events. Unless there is a lot of code, it is probably easier to change the code, rather than change the event chronology.

Specify Compatibility Settings

To specify the compatibility settings, complete the following steps.

For more information, see “VerifierFormLoad” and "FocusChanged" in the *Oracle WebCenter Forms Recognition Scripting User's Guide*.

1. From the **Options** menu, select **Settings**.
2. In the **Settings** dialog box, on the **Compatibility** tab, under **Inheritance Chronology - Event Sequence**, select one of the following options.
    - To set the sequence from parent class to derived class, select **Top Down**.
    - To set the sequence from derived class to parent class, clear **Top Down**.
3. Under the **Advanced Validation Option**, select any of the following options.
    - 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 event when in Verifier Test/Train Modes

Runtime Mode User Interface

Statistic View tab

The Statistic View tab in the upper left pane 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 Tree View tab

The Class Tree View tab displays a tree view of the classification distribution. The view continuously updates at runtime.

Document pane

The upper right pane displays the current document.

Files pane

The lower left pane displays the classification, recognition, and extraction information for the processed files.

Results pane

The lower right pane displays the extraction and reference results.

About the 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.

For more information, see About verification projects.

Switch to Verifier Design Mode

To switch to Verifier Design Mode, complete the following step.

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

About the Verifier Test Mode

The Verifier Test Mode simulates the behavior of Verifier. Use this mode to test the verification with the forms created in Designer's Verifier Design Mode.

For more information, see Verification testing.

Switch to Verifier Test Mode

To switch to Verifier Test Mode, complete the following step.

- On the toolbar, from the Switch to Verifier Design Mode list , select Verifier Test Mode.

What are Workdocs?

WebCenter Forms Recognition creates workdocs from the files provided as input during the initial import step. The OCR step includes a layout analysis that identifies words and blocks within the document. All subsequent workflow steps add their 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 processing of a document is complete when the workdoc is assigned to a document class and all the fields for extraction are filled and valid.

The Document Pane

When you select your document input, the currently selected document displays in the document pane. The pane displays the current page of a file.

The status bar at the bottom displays the total number of documents in the current set and the file name of the current document.

The document navigation bar in the main 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.
4	Displays the number of the current document in the batch. To move to a particular document, enter its number and then press Enter.

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

Button	Description
↕	Adjust the document display to the height of the viewer window.

↔	Adjust the document display to the width of the viewer window.
✥	Adjust the document display to the height or the width of the viewer window so that the entire page displays.
⊕	Zoom in.
⊖	Zoom out.
⏮	Move to the first page of a multi-page document.
◀	Move to the previous page of a multi-page document.
▶	Move to the next page of a multi-page document.
⏭	Move to the last page of a multi-page document.

Switch Between the Workdoc and the Page Definition View

In Definition Mode, you can display either the workdoc or the page definition. By default, the workdoc displays. If workdocs are visible in the field list of Design Mode, the text displays dimmed. When results of the execution display, the foreground displays black. To create reading zones on a document, you need to work with the page definition view. To display the page definition view, complete the following step.

- From the View menu, select Show Page.

Document Processing

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 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

The processing status is shown in the status bar. Note that the initial OCR may take a while.

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.

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.

The Learning Toolbar

Button	Description
❗	Processes the current document. The drop-down menu enables or disables debug mode. This button is available in **Definition Mode** and **Train Mode**.
▶	Processes the next document. This button is available in **Definition Mode**.
⏩	Processes documents in the current set starting with the current document. This button is available in **Definition Mode** and **Runtime Mode**.
⏸	Pauses the processing of documents. This button is available in **Definition Mode** and **Runtime Mode**.
⏹	Stops the processing of documents and clears any results from previous runs. This button is available in **Definition Mode** and **Runtime Mode**.
✚	Adds the current document to the learnset of the selected class. This button is available in **Verifier Train Mode**.
💡	Starts learning for the current project. This button is available in **Definition Mode**, **Train Mode**, and **Verifier Train Mode**.

Compress a Learnset

WebCenter Forms Recognition saves the workdoc files in a compressed format. To update the field identifiers and compress the learnset, complete the following steps.

1. Switch to **Definition Mode**, **Train Mode** or **Verifier Train Mode**.
2. From the **Options** menu, select **Compress Learnset**.
3. In the **Compress Learn Set** dialog box, click **Yes**.

About Highlighting

Highlighting visualizes structures 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 the project uses table analysis, only the best table candidate highlights. Table highlighting is not triggered using toolbar buttons, but is active by default. The behavior varies with the different modes.

- In Definition Mode, table elements highlight during the definition process, and the entire table highlights when a document has been processed.

- In Verifier Test Mode, the extracted table is available and you can use it to control which elements of the document table highlight.

The Highlighting Toolbar

The document highlighting toolbar provides highlighting options for the displayed documents. These buttons act as switches.

Button	Description	View menu item
☐	Displays the document without highlighting.	Highlight Nothing
▤	Highlights all words in the document in turquoise. Words result from the OCR processing. Each character is stored together with its position. To view the OCR result, move the pointer to a word. The recognized word displays as a tooltip.	Highlight Words
🟨	Highlights all blocks in the document in yellow. Blocks are groups of words identified in Definition Mode 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. Move the pointer to a candidate to view the weight. The weight displays as a tooltip.	Highlight Candidates
▤	Highlights all fields in the document. Valid fields highlighted in green, invalid fields highlight in red. Move the pointer to a field to view the name. 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

Print Documents

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

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

What are Project Files?

In Designer, the automatic analysis and processing of incoming documents relies on a group of settings and on trained knowledge. Designer 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 Verifier
- Scripts for document classification, data extraction, data validation, data verification, and document export

Open a Project created with an Earlier Version

Designer can automatically convert existing projects to the current version. The first time you open a project created in an earlier version, all new settings have their default values applied. To review the settings, complete the following steps.

1. Launch Designer.
2. From the **File** menu, select **Load Project**.
3. In the **Oracle WebCenter Forms Recognition Designer – Load Project** dialog box, select a project and then click **Open**.
4. Review and if required, modify the project settings.
5. Repeat the previous steps for all projects.

Create a Project

In Designer, there is only one active project. You do not have to log on when you create a new project. When you save a new project, Designer asks for the Administrator credentials. To create a new project, complete the following steps.

1. Complete one of the following substeps.
    - Launch Designer. This creates a new project automatically.
    - In Designer, from the **File** menu, select **New Project**.
2. From the **File** menu, select **Save Project**.
3. In the **Oracle WebCenter Forms Recognition Designer – Save Project As** dialog box, in the **File** name field, enter the project name and then click **Save**.

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.

Save a Project Without Learn Data

By default, Designer saves the project with learn data. To save the project without learn data, complete the following steps.

1. From the **File** menu, select **Save Project As**.
2. In the **Oracle WebCenter Forms Recognition – Save Project As** dialog box, in the **File name** field, enter the project name.
3. In the **Save as type** field, select **<project name> (SDP)** or **<project name> without learn data** and then click **Save**.

About the Save Compress Feature

Use this feature to save a project file after removing a document class and its associated learnset without causing the project file to grow or remain the same rather than to decrease.

Save 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 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 actively work on the project.

However, “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, use the “Save Project As” option.

Save Compress a Project

To Save Compress a project, complete the following step.

- From the File menu, select Save Compress.

Open a Project

To open a project in Designer, you must be a member of a group that has the ADM role. To open a project, complete the following steps.

1. From the **File** menu, select **Load Project.**
2. In the **Oracle WebCenter Forms Recognition – Load Project** dialog box, select the project file.
3. Complete one of the following substeps.
    - To open the project with learn data, select **Oracle WebCenter Forms Recognition Project (*.sdp)**.
    - To open the project without learn data, select **Oracle WebCenter Forms Recognition Project, skip learn data (*.sdp)**.
4. Click **Open**.

Version Control

Designer provides a version-control feature.

Create a Project Version

To create a new project version, complete the following steps.

1. From the **File** menu, select **Versions**.
2. Optional. In the **Versions** dialog box, in the **History depth** field, enter the maximum number of versions you want to allow for the project.

   Note: If you create more versions than specified in History depth, the new version replaces the oldest version.
3. In the **Versions** dialog box, click the **Save as new** button.
4. In the **Oracle WebCenter Forms Recognition – Create new version** dialog box, enter a description, click **OK** and then click **Close**.

Open a Project Version

To open a project version, complete the following steps.

1. From the **File** menu, select **Versions**.
2. In the **Versions** dialog box, select a version and then click **Open**.

Delete a Project Version

To delete a project version, complete the following steps.

1. From the **File** menu, select **Versions**.
2. In the **Versions** dialog box, select a version and then click **Delete**.

About Client Keys

Client keys are available in the Designer 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.

Display the Client Keys

To display the client keys, complete the following steps.

1. From the **Options** menu, select **Settings**.
2. In the **Settings** dialog box, select the **Client Keys** tab.
3. To display the client-specific values, from the **Active Client** list, select the required client.

Create a Client

By default, each project contains a client named Default. To create an additional client, complete the following steps.

1. From the **Options** menu, select **Settings**.
2. On the **Client Keys** tab, click **Add**.
3. In the **Add Client** dialog box, enter a name for the client and then click **OK**.
4. To select a client, in the **Client Keys** tab, from the **Active Client** list box, select an entry.

Delete a Client

To delete a client, complete the following steps.

1. From the **Options** menu, select **Settings**.
2. On the **Client Keys** tab, from the **Active Client** list, select a client and then click **Remove**.
3. In the **Remove Active Client** dialog box, click **Yes**.

Delete all Clients

To delete all clients, complete the following steps.

1. From the **Options** menu, select **Settings**.
2. On the **Client Keys** tab, click **Clear Settings**.
3. In the **Clear Settings** dialog box, click **Yes**.

Add a Root Key

You can create a hierarchical structure of keys consisting of root and child keys. These keys are available in all clients. To add a root key, complete the following steps.

1. From the **Options** menu, select **Settings**.
2. On the **Client Keys** tab, in the **Key** and **Value** list, right-click and then click **Add Root Key**.
3. In the **Add New Key** dialog box, enter a unique name for the key and then click **OK**.

Add a Child Key

To add a child key, complete the following steps.

1. From the **Options** menu, select **Settings**.
2. On the **Client Keys** tab, in the **Key** and **Value** list, right-click an existing key and then click **Add Child Key**.
3. In the **Add New Key** dialog box, enter a name for the key and then click **OK**.

Delete a Key

To delete a key, complete the following steps.

1. From the **Options** menu, select **Settings**.
2. On the **Client Keys** tab, in the **Key** and **Value** list, right-click an existing key and then click **Remove Key**.
3. In the **Remove Key** dialog box, click **Yes**.

Specify a Value for a Key

For each key, you can specify a value, which the system uses for the selected client.

1. From the **Options** menu, select **Settings**.
2. On the **Client Keys** tab, from the **Active Client** list, select the required client.
3. In the **Key** and **Value** list, select a key and then click the **Value** column for the key.
4. In the **Value** field, enter the value and then press Enter.

About Absolute and Relative Directories in the Project

The project file is portable so that you can easily reuse it. Designer can work either with absolute or relative references to the following directories.

- Export directory
- Learnset directory

Enable Relative References

By default, the project uses absolute references. If you want to copy a project to another directory, use relative references. To enable relative references, complete the following steps.

1. From the **Options** menu, select **Settings**.
2. In the **Settings** dialog box, select **Save all path information relative to the project file** and then click **OK**.

Copy a Project

To copy a project, complete the following steps.

1. Copy the SDP file and, if applicable, the documents from the learnset to the target directory.
2. When using absolute references, complete the following substeps.
    1. From the **Options** menu, select **Settings**.
    2. In the **Settings** dialog box, on the **Train Mode** tab, in the **Base Directory** field, enter or browse to the required directory.
    3. On the **Runtime Mode** tab, in the **Export Directory** field, enter or browse to the required directory and then click **OK**.

About the Project Structure

A standard WebCenter Forms Recognition installation stores the sample projects in the *[Installation path]\Projects* directory. Each project has its own directory structure. The project root directory name corresponds to the project name. Directory structures for projects that do not use associative learning differ from those that use associative learning.

About the Structure of Non-Associative Projects

In non-associative projects, the project file, the .*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 export and learnset directories. The learn directory contains all project learnsets. Each learnset directory contains the images and workdocs in the learnset. The Export directory contains the exported files.

Project Planning

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

About the 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. Although WebCenter Forms Recognition can optimize the OCR result, 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.

About 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 sizes

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

About 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

About 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
- Check Analysis Engine
- Format Analysis Engine
- Table Analysis Engine
- Zone Analysis Engine
- Line Extraction Engine

About Verification

WebCenter Forms Recognition features a dedicated program for quality assurance: Oracle WebCenter Forms Recognition 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.

About 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.

About Document Export

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

For more information, see Specify the Runtime Mode settings and About setting up the document export.

About Document Separation

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

For more information, see “Configure an import instance for document separation” in the *Oracle WebCenter Forms Recognition Runtime Server User's Guide*.

Prepare a Learn Set 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, enter 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, see Fields in the document separation learnset.
    1. InvoiceNumber
    2. Date
    3. Page
    4. PageNumber
    5. 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.

Fields in the Document Separation Learn Set

InvoiceNumber

A principle identifier of a multi-page document. If the invoice number changes between two pages, a new document starts.

Date

Another criterion that can strengthen the determination of a multi-page document.

Page

This field must contain the page keyword, such as “Page” or “Pg.Nr.” to strengthen the page number extraction.

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, you can use the tax number of the vendor 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 start.

Prepare a Project for Document Separation

To prepare a project for document separation, complete the following steps.

1. From the **Options** menu, select **Settings**.
2. In the **Settings** dialog box, on the **Input Mode** tab, select **Save Execution Results to Workdoc** and then click **OK**.
3. Switch to **Definition Mode**.
4. In the left pane, on the **Classes** tab, right-click the project and then click **Show Properties**.
5. In the right pane, on the **Classification** tab, from the **Default Classification Result** list, select a class.
6. In the left pane, on the **Classes** tab, right-click the project and then click **Show Multipage Detection**.
7. In the right pane, on the **General Settings** tab, complete the following substeps.
    1. Select **Activate Multipage Detection**.
    2. From the **Choose Default** list, to specify the detection starting point, select **First Page** or **Following Page**.
    3. Optional. To keep the page sizes similar, select **Similar Page Size Required** and in the **Tolerance** box, enter a number or click an arrow.
8. In the right pane, on the **Triton ADS Engine** tab, complete the following substeps.
    1. Select **Activate**.
    2. Adjust the values for **Threshold** and **Distance**.
    3. In the **ADS Learn Folder** field, enter or browse to the learnset directory.
9. From the **File** menu, select **Save Project**.

About Multipage Detection

Multipage detection uses phrases to find documents that belong together and makes multiple page TIFF 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.

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

For more information, see “Configure an instance for multipage detection” in the *Oracle WebCenter Forms Recognition Runtime Server User's Guide*.

About Classification

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

The confidence value returned by a classification engine indicates the degree of similarity between a document to process 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 adjust the threshold values and the evaluation algorithms, the default settings are applicable for most cases.

Project Classification Properties

- **Threshold**: WebCenter Forms Recognition can classify a document if its confidence value with respect to a given class exceeds the specified threshold.

  The default threshold is 70 percent.

- **Distance**: 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. This requires a certain difference in confidence between the winning class and the second-best competitor.

  By default, the distance is 20 percent.

- **NO Threshold**: Some classification engines provide the possibility to specify this value, which defines that a document cannot belong to a specific class if the class weight is below this value.

Notes

- If you applied multiple classification methods to a given class, WebCenter Forms Recognition computes a combined result.

- Ensure to scan all documents you want to process within the same project to the same resolution.

  If you want to use the WebCenter Forms Recognition Table Extraction engine, the resolution of all documents must be 300 DPI.

About Advanced Classification

In Designer you can configure the classification matrix for Verifier users. This feature suggests several probable classes if Designer cannot determine the class result by 100 percent. If more than one class is in the list, Designer supplements the class entry by a value representing the determination probability. The class with the highest probability displays on top of the list.

Enable Advanced Classification

To enable advanced classification for Verifier, complete the following steps.

1. From the **Options** menu, select **Settings**.
2. In the **Settings** dialog box, on the **Verifier Mode** tab, select **Enable advanced classification in Verifier** and then click **OK**.

Prepare Sample and Test Documents

Before you begin designing your project, you need enough documents to prepare and test your project. To prepare the sample and test documents, complete the following steps.

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

Prepare the Project

To prepare a project for classification, complete the following steps.

1. From the **Options** menu, select **Settings**.
2. In the **Settings** dialog box, on the **Runtime Mode** tab, in the **Export Directory** field, enter or browse to the directory in which you want to save the processing results.
3. On the **Train Mode** tab, in the **Base Directory** field, enter or browse to the directory in which you want to copy your sample documents.
4. On the **Input Mode** tab, select a job or specify the batch root directory.
5. From the **File** menu, select **Save**.

About Classes

Designer enables you to create multi-level hierarchical classification trees 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.

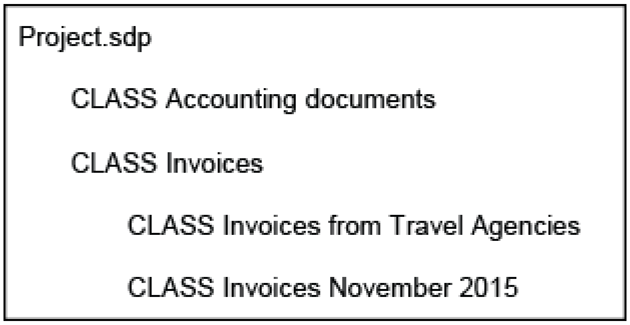
When you create a class hierarchy, consider the following suggestions.

- Create derived classes as 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.

Example

A classification tree as shown below does not yield satisfactory results.



Project.sdp
- CLASS Accounting documents
- CLASS Invoices
  - CLASS Invoices from Travel Agencies
  - CLASS Invoices November 2015

Such a classification tree is insufficient because the **Invoices** class includes two derived classes that are not parallel, **Invoices from Travel Agencies** and **Invoices November 2015**. It is conceivable that **Invoices November 2015** would also include invoices that came from travel agencies and therefore would be part of the **Travel Agencies** class.

A better classification tree might consist of **Invoices from Travel Agencies** and **Invoices from Mechanics**, or **Invoices November 2015** and **Invoices December 2015**.

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, in the left pane, on the **Classes** tab, right-click the project and then click **Insert Base DocClass**.
3. In the **Create New Base DocClass** dialog box, enter a name for the class and then click **OK**.

   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, see About display names.
4. Click **OK** to confirm. The new class is inserted below the root class in alphabetical order.
5. Repeat this process until all base classes are created.
6. To add a derived class, right-click a class and then click **Insert Derived DocClass**.
7. In the **Create New Derived DocClass** dialog box, enter a name for the class and then click **OK.**

About Display Names

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

The feature is important for some supervised learning configurations and usage of the learnset manager tool, in case the reviewed system class names do not have meaningful names.

Both Verifier and Designer use the classes' display names in the list view items and the status bar indicator.

- The Fields view of the Definition Mode of Designer, including both status bar indicators of the document class of the opened document it is currently classified to.
- Normal Train and Verifier Test modes, 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.
- Manual Classification view of Verifier, including the list items of the available classes' list and the currently selected class, as shown in the following figure.
- Global learnset Browsing mode of learnset Manager Tool that you can launch from Verifier.

Note: The other class view modes of remain unaffected.

Specify a Display Name for a class

To specify a display name for a class, complete the following steps.

1. On the toolbar, click the **Switch to Definition Mode** button.
2. Right-click the class and then click **Show Properties**.
3. In the right pane, on the **Document Class** tab, in the **Display Name** field, enter the display name for the class.
4. In the left pane, select the project.
5. In the right pane, on the **Advanced** tab, select **Show custom class names**.

   Note: By default, this option is enabled.
6. From the **File** menu, select **Save Project**.

Delete a Class

If you delete a class, you lose all trained knowledge for the project. To delete a class, complete the following steps.

1. Switch to **Definition Mode**.
2. In the left pane, on the **Classes** tab, right click the class and then click **Delete DocClass**.
3. From the **Options** menu, clear **Incremental Learning**.
4. On the **Toolbar**, click the **Learn documents** button.

Rename a Class

If you rename a class, you lose the learnset for this class and all trained knowledge for the project. To rename a class, complete the following steps.

1. Switch to **Definition Mode**.
2. In the left pane, on the **Classes** tab, right click the class and then click **Rename DocClass**.
3. In the **Rename DocClass** dialog box, enter a new name for the class and then click **OK**.
4. Create a new learnset for the class and repeat the training.

Move a Class

You can move classes within the classification tree to change the class hierarchy. If you move a class, you lose all trained knowledge for the project. If you used inherited field properties, check whether your settings are still consistent.

To move a class, complete the following steps.

1. Switch to **Definition Mode**.
2. In the left pane, on the **Classes** tab, drag the class to the new place in the hierarchy.
3. In the right pane, verify the class properties.
4. Repeat the training.

About Classification Methods

Once you have created classes, specify the classification methods for each class.

Try a single classification method first. We recommend to use the WebCenter Forms Recognition Classify Engine as the primary method. If you need to improve the results, you can use complementary methods. You can freely combine the classification methods.

Select Classification Methods

To select one or more classification methods, 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, select the methods that you want to use for the selected class.
4. Repeat the previous steps for all classes.

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.

Example using the maximum method

Class	Result
Class A	88
Class B	67
Class C	63

The threshold is set to 70 and the distance is set to 20.

Result

- Class A is the best class. The confidence for class A is above the threshold.
- Class B is the second-best class. The distance to class A is 21. This is just above the required distance.
- The document is assigned to class A.

Example using the average method

Class	Result
Class A	78
Class B	63.5
Class C	62.5

The threshold is set to 70 and the distance is set to 20.

Result

- Class A is the best class. The confidence for class A is above the threshold.
- Class B is the second-best class. The distance to class A is 14.5. This is just below the required distance.
- The document cannot be classified using standard classification.

Example using the weighted distance method

Class	Result
Class A	68
Class B	18
Class C	8

The threshold is set to 70 and the distance is set to 20.

Result

- Class A is the best class. 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 probably have to work with smaller thresholds to obtain a satisfactory fraction of classified documents.

About Project-Level Parent Classification

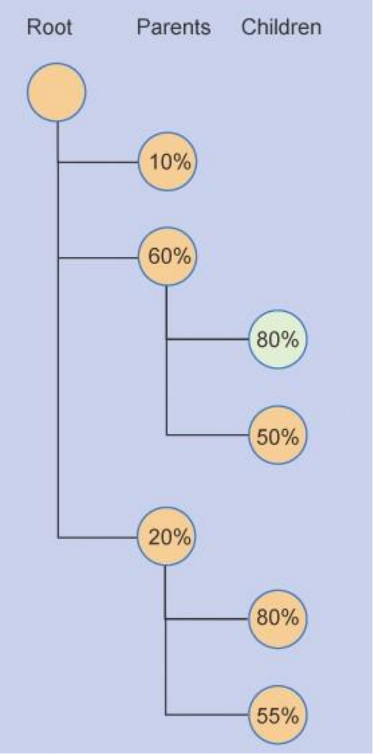
So far, no hierarchical elements were taken into account for classification. However, if classification is not possible using the flat mechanism, hierarchical elements may resolve the conflict.

One of the available hierarchical methods is parent classification. If parent and child classes compete for the same documents, the documents are rather 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 the following additional values defined at the project level.

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

In the following figure, classification occurs to the medium branch, and the child class with 80 percent confidence wins.



For more information about hierarchical classification elements, see [About class-level subtree classification]().

About Project-Level Default Classes

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

During processing, 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 is used if classification fails. You can use this feature also to create applications without classification, but with extraction.

Modify Project-Level Settings

To modify the project-level settings, 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 **Default Classification Result** list, select the required default class.
4. In the right pane, on the **Project** tab, complete any of the following substeps.
    - To change the algorithm to obtain combined results for given classes, select an entry from the **Classification Interpretation** list.
    - To change **Threshold** or **Distance for Standard Classification**, use the sliders or enter values into the corresponding text boxes.
    - To change **Threshold** or **Distance for Parent Classification**, use the sliders or enter values into the corresponding text boxes.

Classification Interpretation

Assume there are given classes, a given document, and several classification methods that return confidence levels. It is unlikely that all methods return the same confidence level. To get the best bet on what the confidence really is, WebCenter Forms Recognition computes a combined result using one of the following methods.

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%

The average method

The average 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%

The weighted distance method

The maximum of all results for all classes is taken as a reference. For each result, WebCenter Forms Recognition calculates the distance to this reference. 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 – (28-27)

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.

About Class-Level Subtree Classification

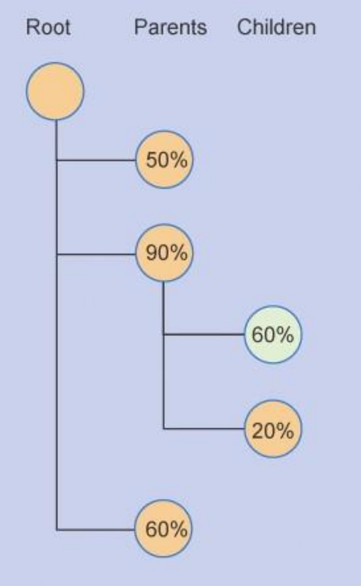
Subtree classification is a means to push classification from parent to child classes by defining reduced requirements for classification into the subtree.

Subtree classification uses the following additional values defined at the class level:

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

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

In the example below, without subtree classification, the parent class with 90 percent confidence would win. With subtree classification, classification would lead to the medium branch, and the child class with 60 percent confidence would win.



About Class-Level Redirection

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.

Modify the Class-Level Settings

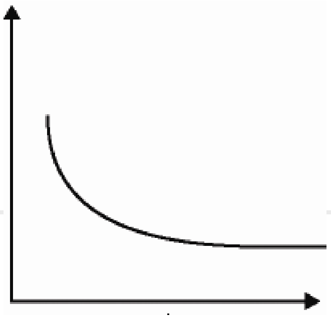
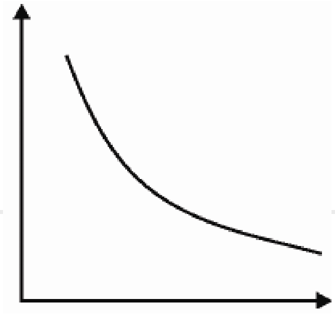
To modify the class-level settings, complete the following steps.

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 **Brainware Classify Engine**.
4. On the **Brainware Classify Engine** tab, select a view.
5. Optional. To use a different display name for the class, on the **Document Class** tab, in the **Display Name** field, enter a name.
6. Optional. To exclude the class as classification result, on the **Document Class** tab, clear the **This DocClass is a valid Classification Result** check box.
7. Optional. To hide the class in **Verifier**, clear the **Visible in correction** check box.

   Note: See also Test the classes and their assigned forms.
8. Optional. To push the classification to the children of this class, select the **Force Subtree Classification** check box and modify the values for **Threshold** and **Distance** under **Subtree classification**.
9. Optional. To redirect the classification result, from the **Redirection Classification Result** list, select a class.
10. Optional. Select the **Redirect after completion of extraction** check box.
11. To apply a static page count to all documents of this class, select the **static page count** check box and specify the number of pages.
12. Under **Table Training**, select either **Automatically using Learnset** or **Manually**.
13. From the **Classification Field** list, select the field which is used as classification field in Supervised Learning.
14. To set up validation rules, select the **Validation** tab.

About Stretching

Stretching improves the results of the Brainware Classify Engine when combined with other classification engines by increasing the distance in confidence levels between results. Stretching means to increase the distance between results given by the classification engine. The stretching factor can vary from 1.0 to 5.0.

Before stretching	After stretching
	

Result percentage Class	Result percentage Class

Example

If the stretching factor is 2.0, the Brainware Classify Engine’s 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

About Calibration

Calibration fine-tunes the various classification engines. Calibration bases on the calibration factor, which can vary from 0.1 to 1.5.

Classification Learnsets

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

Consider the following suggestions for selecting the sample documents for the learnsets.

- You need a separate learnset for each class.
- You need sample documents that truly represent the topic you want to cover.
- You need at least five sample documents per class. The maximum number of samples per class is limited by memory.
- Use single topic documents.
- Use unique documents for each 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.
- The number of samples needed to create a good learnset depends on the classification task. The broader a class is supposed to cover, the more samples are required.
- Increase the number of samples if classes are very similar to each other.
- Ideally, use a similar number of samples for each class.

The following classification engines require learnsets

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

Create a Learnset

To create a learnset, complete the following steps.

Prerequisite

Ensure that the “Add trained documents to the learnset” option is active on the Train Mode tab in the Settings dialog box.

1. Switch to **Document Input Selection - Batch**.
2. In the upper left pane, double-click the batch that contains the documents for your learnset.
3. In the lower left pane, select the first document.
4. Switch to **Train Mode**.

   The lower left pane on the **Classes** tab displays the number of documents in the learnset.
5. Complete the following substeps.
   1. If required, use the navigation bar in the toolbar to browse to the document that you want to add to the learnset.
   2. In the left pane, on the **Classes** tab, select a target class from the list in the lower pane.
   3. In the toolbar, click the **Add document to Learnset** button to add the current document to the classification learnset of the selected class.

      The entry in the **Class** column in the upper pane of the **Classes** tab indicates that a document has been added. WebCenter Forms Recognition writes a copy of the document to a protected directory structure that is reserved for learnsets.
   4. Repeat the previous step until you have assigned an adequate number of samples from the current document set.
   5. Repeat the previous steps to add documents for other classes.
6. From the **File** menu, select **Save Project**.

About Learnset Encryption

Use the learnset encryption for secure distribution of learnsets that contain critical data. You can activate the encryption of learnsets for any Designer 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 a 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 the 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, the 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.

Learn Set 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.

Encrypt a Learnset

To encrypt a learnset, complete the following steps.

1. Create a backup of the learnset directory.
2. Switch to **Definition Mode** or **Train Mode**.
3. From the **Options** menu, select **Encrypt Learnset**.
4. In the **Warning** dialog box, click **Yes**.
5. In the **Training Manager** dialog box, click **OK**.
6. Optional. If you added new documents to the learnset and want to encrypt them, repeat the previous steps.

About 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.

Apply a Custom Encryption Key

To create the registry key, complete the following steps.

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 > String Value**.
4. In the **Name** field, enter `EncryptLearnSetKey` and then press Enter.
5. Right-click the **EncryptLearnSetKey** key and then click **Modify**.
6. In the **Edit String** dialog box, in the **Value data** field, enter the custom encryption key and then click **OK**.

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

Correct a Learnset

You can add documents to existing learnsets at any time by performing the steps in Create a learnset. If a document was added to a learnset by mistake, or is assigned to the wrong class, complete the following steps.

1. From the **Switch to Document Input Selection** list, select **Learn Set**.
2. In the upper left pane, select a class to display the documents in its learnset in the lower left pane.
3. In the lower pane, complete any of the following steps.
    - In the lower left pane, select a document to display it in the right pane.
    - To delete a document from the learnset, right-click the document and then click **Remove from learnset**.
    - Drag a document from the learnset to another class.
    - To exclude a document from learning for specified classification methods, clear the respective column.

About Manual Addition of Documents to a Learnset

Although you can add new documents to a learnset at any time, you may find inconsistent results 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 it extracts information.

Without proper learning, the classification does not include the information from the new documents. To remedy this, change the learnset by manually adding a document to it and then changing the working mode. The documents 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, see About learning.

Manually Add a Document to a Learnset

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

Prerequisite Ensure that the “Add trained documents to the learnset” option is active on the

Train Mode tab in the Settings dialog box.

1. From the **Switch to Document Input Selection** list, select **Batch**.
2. In the upper pane, double-click a batch.
3. In the lower pane, select a document.
4. Switch to **Train Mode**.
5. In the left pane, select the **Classes** tab.
6. Complete the following substeps.
    1. Repeat the previous steps for all required documents.
    2. In the lower pane, select a class and navigate to the document you want to add.
    3. On the toolbar, click the **Add document to Learn Set** button to add the current document to the classification learnset of the selected class. Designer creates a copy of the document in a protected directory reserved for learnsets.
7. Switch to another mode to automatically induce learning for the newly added documents.
8. From the **File** menu, select **Save Project**.

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 any of the following changes.

- 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.

Use the Learn Mode

To use the learn mode, complete the following steps.

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

1. Switch to **Definition Mode** or **Train Mode**.
2. Complete one of the following substeps.
    - To effect learning for changes to the learnset only, on the **Options** menu, enable **Incremental Learning**.
    - To effect learning for the entire project, on the **Options** menu, clear **Incremental Learning**.
3. In the toolbar, click the **Learn documents** button.

    Note: Learning may take several minutes.

Check the Learn Status of a Class

Learning always attempts to cover the entire classification tree and omits only classes with insufficient learnsets. 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 right pane, on the **Classification** tab, 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.

About the Brainware Classify 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.

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

Configure the 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, see Restrict the dictionary creation.

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

Restrict the 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, see the *Oracle WebCenter Forms Recognition Runtime Server User's 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 you want to be taken into account.

   Example

   - 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.

Create a View for the Brainware Classify Engine

For instructions on how to assign classes to views, see Modify the class-level settings.

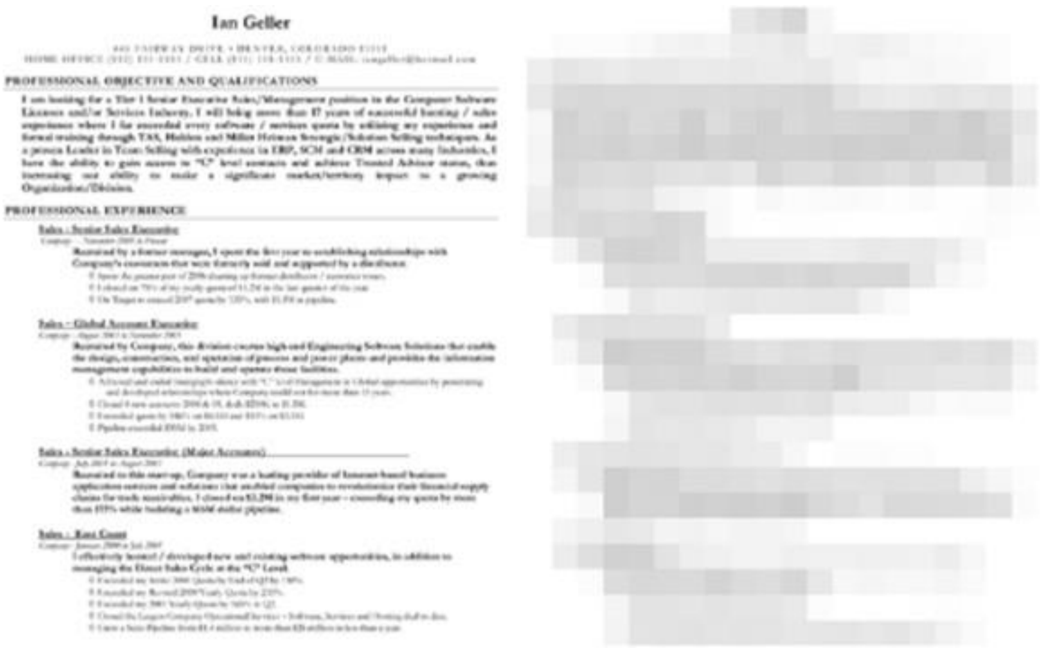
Each project automatically contains a view called *Default View*. By default, all classes use this view. To create a view, complete the following steps.

1. Switch to **Definition Mode**.
2. 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 Classify Engine**.
4. On the **Brainware Classify** tab, click **Insert**.
5. In the **New Brainware View** dialog box, enter a name without spaces for the view and then click **OK**.

About the 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.

Example



Ian Geller

PROFESSIONAL OBJECTIVE AND QUALIFICATIONS

PROFESSIONAL EXPERIENCE

Configure the 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 substeps.
    1. Select **Restrict engine to region**.
    2. Specify the regions you want to take into account.
8. Optional. On the **Algorithm** tab, specify the required resolution.

About the ASSA Classify Engine

Like the Brainware Classify Engine, the ASSA Classify Engine needs a learnset of documents for classification. The difference between the Brainware Engine and the ASSA 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, Brainware Classify Engine generates a dictionary of words. The documents for the search pool are coded with words from this dictionary.

The ASSA engine uses the trigram method to code the search pool. The classification of documents bases 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 engine then searches for that query within the search pool.

Configure the ASSA Classify Engine

To 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 substeps.
   1. Select **Restrict engine to region**.
   2. Select the required pages.
   3. 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.

About the Quality to Speed Relation

This option offers the ability 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.

About the 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.

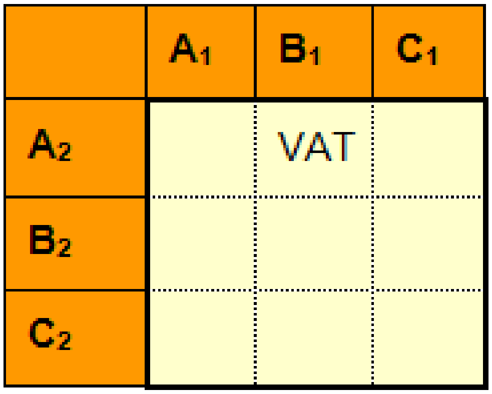
The Brainware Layout Classification engine requires at least two classes defined within a project.

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

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

To create a separate learnset for Brainware Layout Classification, see Create a learnset.
For more information about learning, see About 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 Classify Engine is used for “content” or “type” classification, the purpose of Brainware Layout Classification (BLC) engine is to provide with 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.



	A_1	B_1	C_1
A_2		VAT	
B_2			
C_2			

The BLC engine simply applies normal Brainware classification. However, each word of both learned and extracted documents is not just used as it 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 couple of zones (yellow area is the first page of the document and “VAT” is the word we are looking for):

Now the “VAT” word is going to be extended as “VAT_BBBB_AAAA” identifying that we look for the word “VAT” only the {B1; A2} region of the document. This then simply means that if for the learned layout class X the word “VAT” was mostly located in region {C1; C2} and for the other class Y it was usually {B1; A2} the system will prefer Y while for normal Brainware classification such a difference between X and Y 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.

Configure the 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, see Restrict the dictionary creation.

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

Modify the Extended Settings

To modify the extended settings for the Brainware Layout Classification engine, complete the following steps.

1. On the **Extended** tab, to specify the number of horizontal zones each document is divided into, modify the **Horizontal covering frequency** value.
2. To specify the number of vertical zones each document is divided into, modify the **Vertical covering frequency** value.
3. Complete one of the following substeps.
    - To treat “123” and “987” as the same sequence of digits of the same length, select **Do not differentiate between digits**. Use this option when most numbers on your documents do not have a specific meaning, for example, in the tables of a typical invoice document.
    - If the documents contain specific sequences of digits that can be critical for layout classification, such as bank account information specific to each vendor class, clear **Do not differentiate between digits**.
4. Optional. Select **Attach delimiters to words**.

    Note: In most cases, delimiters such as “,”, “.”, “/”, “;”,are essential to distinguish between words. A US date format, for example, may be better treated as a single differentiating word – “00/00/000” - in the classification’s surrounding compared with 3 “noise” words “00”, “00” and “0000”.

5. Save and relearn the project to apply the new settings.

Create a View for the Brainware Layout Classification

For instructions on how to assign classes to views, see 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, enter a name without spaces for the view and then click **OK**.

About the Forms Classify Engine

The Forms Classify Engine identifies forms or other structured documents that have an identifier of the document class printed on them. If the identifier is placed at a fixed position on the document, reading this zone can quickly provide a classification result. Forms classification is well suited to assigning structured documents to a class and combining several forms in one class. However, the document input is usually too heterogeneous to use forms classification as a stand-alone method.

About 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 you want the project to recognize. In WebCenter Forms Recognition context, three different types of reading zones exist – OCR, OMR and bar codes. However, in forms classification context, you can only use OCR and bar code zones.

- 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 bar code recognition. The number of strings represented by the bar code is the result of this process.

For more information about OMR zones, see About OMR zones.

Configure the 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.

Create Reading Zones

You can use more than one reading zone on each form. To create a reading zone, complete the following steps.

Prerequisite You must have set up a classification scheme.

Note: For more information about creating and editing reading zones, see About reading zones and About the Zone Analysis Engine.

1. Load a document that belongs to the selected class into the viewer.
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. From the **View** menu, select **Show Page**.
6. On the document toolbar, click one of the following buttons to specify the zone type.
    - To activate the selection tool, click the **Select object** button.
    - 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 bar code reading zone, click the **Insert barcode zone** button
    - To create an anchor, click the **Insert anchor** button.
7. In the document, click and drag to create the reading zone around the form identifier.

   Note: Make the reading zone minimally larger than the identifier. 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.

8. In the left pane, select the project.
9. In the right pane, on the **Classification** tab, select the **Forms Classify Engine**.
10. 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.

Configure the Form Identifiers

To configure the form identifiers, complete the following steps. These form identifiers apply only to 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 substeps.
    1. In the **Zone Name** column, select a zone from the list.
    2. In the **Compare Text** column, enter the string that identifies your form.

       Note: This string is compared with the recognition result from the reading zone.

3. In the **Page Number** column, select **1** if the reading zone is on the last page, or **LastPage** if the reading zone is on the last page

5. Repeat the previous step for all reading zones you want to take into account.

Specify the Compare Method

By default, WebCenter Forms Recognition compares the recognition results and form identifiers using an error-tolerant algorithm. To specify the comparison method in more detail, complete the following steps.

1. On the **Forms Classify** tab, select the required row.
2. Click the **Advanced Settings** button.
3. In the **Advanced Settings** dialog box, complete the following substeps.
    1. Optional. In the **Compare Text** field, modify the string.

Note: This might be necessary depending on the selected comparison method.

2. From the **Compare Method** list, select one of the methods.

   Note: For more information, see Compare Methods.

4. If you want to assign several forms with similar identifiers or several versions of the same form to the same class, it may be useful to ignore leading or trailing characters of the recognition string. Complete any of the following substeps.
   - In the **Ignore first characters** box, enter a number or click an arrow.
   - In the **Ignore last characters** box, enter a number or click an arrow.
5. To perform the comparison neglecting leading or trailing characters of the comparison string, in the **Max. compare distance** box, enter a value or click an arrow. The compare distance computes as follows. Compare Distance = 1.00 - Number of Characters (Recognition String) / Number of Characters (Comparison String).

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

6. In some cases, you may need multiple reading zones to obtain a reliable classification. In this case, you need to specify how the classification result is evaluated. To set the number of zones that must match, complete one of the following substeps.
   - On the **Forms Classify** tab, select **Match all zones**.
   - On the **Forms Classify** tab, clear **Match all zones** and then, to specify the number of required zones, in the **Required number of** box, enter a value or click an arrow.

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.

Expression

A method that returns a match if the recognition result corresponds to a possibly complex format pattern that needs to be entered into the Compare Text field.

Simple Expression

A method that returns a match if the recognition result corresponds to a simple format pattern that needs to be entered into the Compare Text field.

Colors in the Compare Text Column

On the Forms Classify tab, the strings in the Compare Text column display in different colors to indicate the selected method.

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

About 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 set it up in the project page to read the zone globally.

Before you share OCR zones, make sure to have at least one class set up with forms analysis.

Set up 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.

Prerequisite: You must have set up at least one class with forms analysis.

1. In the left pane, on the **Classes** tab, select the project.
2. Decrease the size of the **All** zone to a small rectangle.
3. Set up a shared OCR zone by completing the steps in Create reading zones.

About the Phrase Classify Engine

The Phrase Classify Engine classifies documents by comparing the document's OCR result with a list of allowed and forbidden words and phrases. Within certain error limits, the phrases or words must literally occur within the documents.

The engine compares the document's OCR result with the defined lists and determines a phrase weighting.

The engine determines a phrase weight that is used for the class, with the threshold and distance settings of the project, against the weighting of the classification results of other classes and other engines.

This engine is a 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 and maintenance of a potentially complex set of rules. Typically, you use the Phrase Classify Engine in conjunction with the Brainware Classify Engine.

Note: The Phrase Classify Engine cannot be trained, so you do not need to add sample documents to the learnset.

About Search Phrases and Special Characters

A search phrase consists of one or more search words. 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. You cannot search for these wildcard characters explicitly.

- * represents zero or more characters within a single word.
- ? represents a single character, but not a space.
- ??, when preceded and followed by spaces, represents a word. If ?? is included within a search word, it represents two characters.

Leading or trailing wildcard characters

Leading or trailing wildcard characters on the search word are not meaningful and will be removed or ignored.

Example: The search word *read* matches any word that contains *read*, such as *bread*, *read*, *reader*, or *breading*. Search words that include leading or trailing wildcard characters, such as *?read* or *read** do not result in any additional matches.

The wildcard word *'??'*

The wildcard word *'??'* can be entered manually, or can be configured through phrase settings by setting a **Maximum distance to next word**. This specifies the possible distance between this search word and the next word.

Example: A value of 1 means the search words must be in adjacent words, while a value of 6 would allow for up to 5 words to appear between this and the next search word.

For more information, see Phrase Settings.

About Candidates

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

When a search phrase consists of a single search word, the engine identifies any word in a document that contains the search word as a candidate. The candidates for the selected search phrase display 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 you set the significance of a search phrase to 'forbidden' and WebCenter Forms Recognition finds at least one reliable candidate, candidates do not display highlighted.

About the Phrase Weight

For each candidate, the engine calculates a weight based on the significance and strength settings for the search words, 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 each search phrase does not affect the phrase weight, though it is taken into account when using the Phrase Classify Engine for classification.

This phrase weight displays as a percentage value between 1 and 100 in the tooltip for each highlighted candidate in the document. However, in the case of a "forbidden" search phrase for which all candidates are questionable, such as errors, the phrase weight may be less than zero.

About the Classification Value

When the classification step uses the Phrase Classify Engine, the output bases 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. If the sum is greater than 100, the classification value is set to 100.

Configure the Phrase Classify Engine

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. On the **Phrase Classify** tab, modify the Distance and parameters.
7. Optional. On the **Regions** tab, to restrict the classification to certain areas within the documents, complete the following substeps.
    1. Select **Restrict engine to region**.
    2. Select the required pages.
    3. For each selected page, specify the percentage that you want to take into account.

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.

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

Prerequisite: You must have set up a classification scheme.

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, enter a keyword or a phrase and then click **Ins**.

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

5. Optional. To add a word or a phrase from the document, complete the following substeps.
   1. On the toolbar, click the **Highlight Words** button.
   2. In the document, click a word to copy it into the **New Phrase** field.
   3. Optional. To build a phrase, click another word.
   4. 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 substeps.
   1. On the toolbar, click the **Highlight nothing** button.
   2. Select the phrase from the list of active phrases.

      Note: Each occurrence is highlighted in green.

   3. Optional. To view the number of typing errors and the phrase weight, point to the highlighted phrase.
8. To specify the phrase settings, complete the following substeps.
   1. Select a phrase in the **Active Phrases** column and then click **Edit**.
   2. In the **Phrase Settings** dialog box, modify the settings as required. The dialog box provides a separate tab for each word in the phrase.

      For more information, see Phrase Settings.

Phrase Settings

Tolerance

Maximum number of not matching characters allowed for this word.

Example: The phrase contains the word “invoice” and the number of not matching characters is set to 2.

Result: Words such as “invoke,” “involve,” and “invite” meet the requirements.

Word strength

The stronger the word, the more typing errors can occur until the word loses its meaning. A maximum strength setting stands for a word that retains full significance with the maximum number of permissible typing errors.

Accept only case-sensitive matching

Select this option to enforce case-sensitivity.

Relative significance of word

Use this option to set the proportional significance of the current word within the phrase.

Example: You can reduce the relative significance of the word “with” in the phrase “with kind regards,”.

Maximum distance to next word

Use this setting to specify whether your phrase is still recognized if there are additional words between the current one and the next one specified.

Example: You can use a single definition for “please pay by” and “please pay this amount by” with a maximum distance of 3. The distance for consecutive words is 1.

Significance

This setting affects the entire phrase. The significance may vary between absolute (a document belongs for sure in this class if the phrase is found) and forbidden (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 image.

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

Image	Description
Degressive	The sum of all candidate weights. **Note:** For any search phrase with one or more candidates, the effective weight of each subsequent candidate is reduced proportionally. For example, a second candidate counts one half, a third candidate counts 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.

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 substeps.
    - 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.

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**.

Export Phrases

You can export phrases to a file and then import them in another project. 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, click **Export**.
4. In the **Save As** dialog box, in the **File name** field, enter a file name and then click **Save**.

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, click **Import**.
4. In the **Open** dialog box, select an EXP file and then click **Open**.

About the Language Classify Engine

The Language Classify Engine automates the classification of processed documents by language. The engine contains predefined and pretrained 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 it then uses for classification purposes. The pre-trained version of this pool is distributed along with the 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 uses 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 non-western languages, complete the steps in Activate multi-byte encoding for CJKT languages.

Configure the 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 substeps.
    - Select **Application path**: Stores the language classification learnset in *[Installation directory]\Oracle WebCenter Forms Recognition\bin\Langpool* and the language source files in *[Installation directory]\Oracle WebCenter Forms Recognition\bin\Langpool\Source*.
    - Select **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, manually create the following directories.
    - *\LangPool*
    - *\LangPool\Source*

    Note: For more information, see About new languages in the standard learnset.

    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)
9. To regenerate the entire ASSA classification learnset, click the **Create New Learn Set** 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.

Create a 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.

About New Languages in the Standard Learnset

When generating a new learnset, the engine automatically searches for all available LNG files in the *\LangPool\Source* directory.

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

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. From the **File** menu, select **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 *\LangPool\Source* directory.

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 Learn Set**.

Exclude a Language from the Learn Set

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 Learn Set**.

Remove a Language from the Learn Set

To remove the language from the list of available languages, complete the following steps.

1. In the *\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 Learn Set**.

Modify an Existing Language Learn Set

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

1. Open the LNG file in **Microsoft Word**.
2. Modify the file.
3. Save the file as Unicode file.
4. Switch to **Definition Mode**.
5. In the left pane, on the **Classes** tab, right-click the project and then click **Show Properties**.
6. In the right pane, on the **Classification** tab, click **Create New Learnset**.

Check the 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**.

About the 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.

Configure the Image Size Classification Engine

To use image size classification, you must specify the document size for the corresponding class. You can also specify a size range.

To configure the Image Size Classification engine, complete the following steps. **Prerequisite:** You must have set up a classification scheme.

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 **Image Size Classification**.
4. On the **Image Size** tab, from the **Minimum Size** list, select a size.
5. From the **Maximum Size** list, select a size.
6. Under **Orientation**, select the page orientation.
7. Under **Tolerance**, specify the error tolerance that the engine uses when comparing sizes.

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

About the Template Classify Engine

The Brainware Layout Classification engine has replaced the Template Classify Engine. For more information, see About the Brainware Layout Classification Engine.

Engine-Level Absolute Results

For a certain class and document, each classification engine may return different result types. The following result types are possible.

Confidence

This document belongs to a specified class with a probability given in percent.

Yes

This document definitely belongs to this class.

Maybe

This document might belong to this class.

No

This document does definitely not belong to this class.

This method is not applied.

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.

Engine	Result	Default or Configured
Brainware Classify	Confidence Yes No	Default Configured Configured
ASSA Classify	Confidence Yes No	Default Configured Configured
Phrase Classify	Confidence Yes No	Default Configured Configured
Image Size Classification	Maybe No	Default Default
Forms Classify	Yes Maybe Confidence	Default Default Configured

Layout Classify	Confidence Yes No	Default Configured Configured
Brainware Layout Classification	Maybe No Confidence	Configured Configured Default
Language Classify	Maybe No Confidence	Configured Configured 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 is Yes. Classification of the document stops at this stage. All other classes are excluded automatically.

If a method returns No for a class, the combined result is No. Within the current branch, classification of the document stops at this stage: All child classes are excluded automatically.

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

About Classification Testing

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

Test the Classification in Definition Mode and Train Mode

To test the classification and to determine target classes and confidence values, complete one of the following steps.

Prerequisites: You must have an active document set and a trained classification scheme.

- To test the classification only, exclude the data extraction from the runtime functionality in the project settings.
- To process the current document, from the Classify / analyze the current document ❗ list, select Execute.
- To process and export the current document, from the 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.
- Click the Process next document ▶ button.

  Note: This button is available in Definition Mode only.
- Click the Process documents starting with the current one ⏩ button.

Note: This button is available in Definition Mode only.

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**.

Classification Results

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

- Unprocessed documents: Not Classified
- If the classification could not assign the document: Uncertain Classification
- For successfully classified documents: The proposed class

Display the Classification Matrix

To display the classification matrix for the previously processed document, complete the following step.

- On the toolbar, click the Show/Hide Classification Matrix ▦ button.

Test the Classification in Runtime Mode

To test the classification and to determine target classes and confidence values, complete one of the following steps.

- Click the Process next document ▶ button.
- Click the Process documents starting with the current one ⏩ button.
- Optional. To stop the processing and clear any results from previous runs, click the ■ Stop / Clear button.

About the Classification Results in Runtime Mode

The upper left pane displays summary results for all classes. The Statistic View tab displays classification and extraction results. The Class Tree View tab displays a tree view of the classification distribution. Both views update continuously at runtime.

The Statistic View Tab

Classification-related columns on the Statistic View tab

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 you stop processing.

Classification Level

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.

Classify 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.

Extraction-related columns in the lower left pane

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.

Create a Reference File

The Statistic View tab enables you to compare the classification results with results stored as reference files.

- To create a reference file, complete one of the following actions.

Situation	Steps
Enable creation of the default export file.	1. From the Options menu, select 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.	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, enter a file name and then click Save.

Modify a Reference File

You can modify reference files as well as the default export file, or you can create a reference file from scratch. Each line in the file represents the results of one document. To modify a reference file, complete the following steps.

1. Open an existing REF file or create a new file with a text editor.
2. Add the required lines to the file according to the following structure.

   <file name><TAB><class><TAB><field 1><TAB>…<TAB><field n>

Example `CaptureInvoice.tif<TAB>Invoices<TAB>$5.489,98<TAB>10324`

3. Save and close the file.

Load a Reference File

To load a reference file, complete the following steps.

1. On the **Statistic View** tab, right-click a class and then click **Load Reference**.
2. In the **Open Reference File** dialog box, select a REF file and then click **Open**.

Compare Classification Results

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

1. On the **Statistic 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.

The Class Tree View Tab

The Class Tree View tab displays the distribution of class assignments.

If you click a class in either the Statistic View or the Class Tree View, see 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

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.

Classification Problems

Most classification problems are due to one of the following situations.

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

Resolve OCR Problems

To identify and resolve OCR problems, complete the following steps.

1. On the toolbar, click one of the highlighting buttons.
2. In case of OCR problems, complete any of the following substeps.
    - Improve the quality of the images, for example ensure to scan the document with a resolution of 300 DPI or review your scanner settings.
    - Adjust the default OCR settings. For more information, see Specify the OCR settings.

      Note: Changing the OCR settings on project level is a severe intervention. Exclude other reasons before you do this.

Resolve Learn Set 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.

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 engine that causes the problem.
3. Determine whether classification fails because the confidence is too low or whether the distance is too low.
4. If classification assigns documents to the wrong class, modify the learnset or the parameters for the engine that returns the wrong result.
5. If too many documents are not classified at all, consider adding another classification method, or provide more samples for the learnset.
6. If the previous steps did not help, you may have to change the settings for classification evaluation.

Recognition

WebCenter Forms Recognition provides engines for OCR, OMR, and bar code recognition.

Recognition performs on demand during the classification and extraction steps. You can perform the recognition as often as required. The system stores the result in the workdoc.

Each engine has default parameters that are applicable in most cases. However, you can customize the recognition settings according to your needs on several levels for each processing step that involves recognition.

OCR (Optical Character Recognition)

Use OCR to provide text input for the classification step. In the extraction step, you can use any of the recognition techniques.

WebCenter Forms Recognition features the following OCR engines.

- FineReader 11
- FineReader 10
- Kadmos5 OCR/ICR Engine

OMR (Optical Mark Recognition)

WebCenter Forms Recognition features the following OMR engine.

- Cairo OMR Engine

Barcode Recognition

WebCenter Forms Recognition features the following OCR engines.

- Cleqs Barcode Engine
- FineReader 11 Barcode Engine
- QualitySoft Barcode Engine

Note: We recommend handling Barcodes which encode multiple words using Zone Analysis Engine and not document recognition. Document recognition assigns the same position value to all words, which can result in incorrect word order during manual extraction.

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.

Notes

- 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.

About FineReader Languages

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

For information about the available languages, see “OCR Engine Languages” in the *Oracle 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:;<=>?@ABCDEFGHIJKLMNOPQRSTUVWXYZ  [\]_`
`abcdefghijklmnopqrstuvwxyz{|}~£¥§©«¬®°±»—„•€™`

Configure the FineReader OCR Engine

To configure the FineReader11 OCR engine, complete the following steps.

1. In the **Field OCR Settings Properties** dialog box, under **Recognition Type**, select **OCR**.
2. From the **Available Engines** list, select **FineReader10** or **FineReader11**.
3. Below **Recognition Trade Off**, depending on whether performance or OCR accuracy is more important, select one of the following settings.
    - To decrease the recognition error rate accepting a higher processing time, select **Most Accurate**. This is the default mode.
    - To perform the recognition between **Most Accurate** and **Fast**, select **Balanced**.
    - To perform a fast recognition accepting less accuracy, select **Fast**.
4. Below **Rotation**, select any of the following options.
    - To automatically detect the page orientation, select **Detect Orientation**.
    - To automatically rotate the document so that the text is vertically aligned, select **Deskew**.
5. Below **Enhancement**, select any of the following options.
    - To remove all horizontal or vertical internal lines, select **Remove Lines**.

- To removes background noise from the image for recognition, select **Remove Texture**.
- To remove speckles from the documents, select **Despeckle**.

  Note: Speckles are a group of black pixels surrounded by white pixels or vice versa.

6. Click the **Recognition** tab and select any of the following options.
   - To automatically detect how the documents have been printed, select **Automatic**:

     Use this option unless your document input is very homogeneous and from a dot-matrix printer or a typewriter.

   - For documents printed with dot-matrix printers, select **Matrix**.
   - For documents created with typewriters, select **Typewriter**.
   - For documents with OCR-A and OCR-B fonts, select **OCR-A / OCR-B**.

     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.

   - For documents using MICR E13 B fonts, such as checks, select **MICR E13 B**.
   - To improve the recognition of text specific to sales receipts and invoices on low quality documents, where the OCR process focuses on tidying up the noises around the actual text, select **Receipt**.
     Note: This setting is only available in FineReader 11.
   - To recognize characters or symbols written slightly above right or left to another character, such as "™" in "Company Product™", select **Allow Superscript**.
   - To recognize characters written slightly below another character, such as the "2" in the chemical formula "H2O", select **Allow Subscript**.
   - To improve single digit columns recognition when using CJKT languages, select **Small Digit for Asian Enh.**.

     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, see About non-western languages.

7. Optional. To add a language, click the **Language** tab and complete the following substeps.
   1. 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: The FineReader engines process the recognition with the languages in the **Used** list.

FineReader11 Barcode Engine

The FineReader11 Barcode engine can recognize various bar codes.

For a list of supported bar codes, see "Supported Bar Codes" in the *Oracle WebCenter Forms Recognition Technical Specifications*.

Configure the FineReader11 Barcode Engine

To configure the FineReader11 Barcode engine, complete the following steps.

1. In the **Field OCR Settings Properties** dialog box, under **Recognition Type**, select **Barcode**.
2. From the **Available Engines** list, select **FineReader11 Barcode**.
3. On the **Barcode Types** tab, select the required bar codes.
4. On the **Extended** tab, complete the following substeps as required.
    1. Use the slider to modify the **Confidence Threshold** value.

       Note: The bar code characters confidence must exceed the configured threshold.
    2. Select one or more **Orientation** values.

       Clearing a value causes the engine to ignore bar codes with that angle during processing.
    3. To enable the recognition of low bar codes, enter a value in the **Min. Height** field.

       This property defines the minimum acceptable height of the bar code in relation to the average letter height.

       To allow the engine to automatically adjust this setting, use the default value `-1`.
    4. If the bar code to be recognized has a checksum, select **Validate Checksum**.

       Note: This setting only applies to **Code 39, Interleaved 2 of 5, Codabar, and Matrix 2 of 5** bar codes.
    5. To let the engine check more options of bar code positions, select **Advanced Extraction**.

       Note: This setting reduces the processing speed and only applies to **2D** bar codes.
    6. If the **Code 39** bar code to be recognized does not have the asterisks as start and stop symbol, select **Without Asterisk**.

       Note: This setting only applies to **Code 39** bar codes.
    7. Click **Apply** and then click **OK**.

QualitySoft Barcode Engine

The QualitySoft Barcode Engine is WebCenter Forms Recognition's default bar code recognition engine. The engine includes the following features.

- Reads all common bar code types
- Recognizes multiple bar codes on a single page
- Reads bar codes oriented at any angle

For a list of supported bar codes, see "Supported Bar Codes" in the *Oracle WebCenter Forms Recognition Technical Specifications Guide*.

Configure the QualitySoft Barcode Engine

To configure the QualitiySoft Barcode engine, complete the following steps.

1. In the **Field OCR Settings Properties** dialog box, under **Recognition Type**, select **Barcode.**
2. From the **Available Engines** list, select **QualitySoft Barcode**.
3. On the **General** tab, complete the following substeps as required.
    - To specify the maximum number of bar codes to recognize within the reading

zone, modify the value in the **Max. number of barcodes** field.

Note: The default value is 10. Reducing the number increases the recognition speed.

- To specify the bar code minimum length in characters, modify the value in the **Min. number of chars** field.
- To specify the bar code maximum length in characters, modify the value in the **Max. number of chars** field.
- Select one or more **Orientation** values and complement the **Orientation** settings with the **max. Skew** settings to get most fast and accurate results.

  Note: Clearing an **Orientation** value causes the engine to ignore bar codes with that angle during processing.
- To triggers an additional bar code detection algorithm which improves the bar code detection, select **Intensive Search**.

  Note: This setting reduces the processing speed.
- To define the maximum angle deviation from the configured main orientations, select one of the **max. Skew** values.

  Note: Setting a small value increases recognition speed.

4. On the **Types** tab, select the required bar codes.

   Note: Excluding bar code types which your documents don’t contain increases the recognition speed.
5. Click **Apply** and then click **OK**.

Cleqs Barcode Engine

The Cleqs Barcode Engine engine can recognize various bar codes.

For a list of supported bar codes, see "Supported Bar Codes" in the *Oracle WebCenter Forms Recognition Technical Specifications*.

Configure the Cleqs Barcode Engine

To configure the Cleqs Barcode engine, complete the following steps.

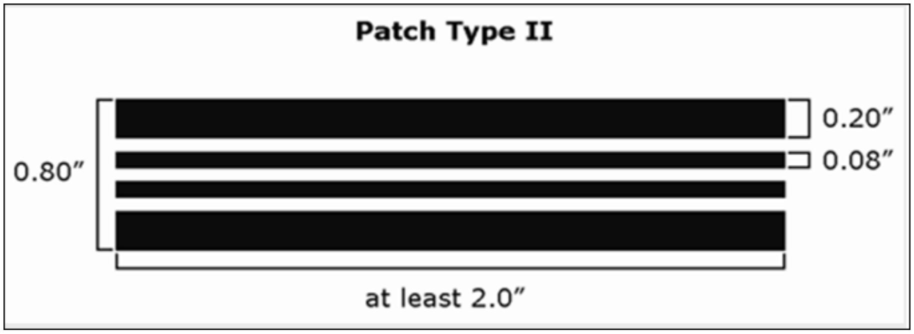
1. In the **Field OCR Settings Properties** dialog box, under **Recognition Type**, select **Barcode.**
2. From the **Available Engines** list, select **Cleqs Barcode**.
3. On the **General** tab, complete the following substeps as required.
   - Use the slider to modify the **Confidence Threshold** value.
   - To specify the maximum number of bar codes to recognize within the reading zone, modify the value in the **Max. number of Barcodes** field.

     Note: The default value is 10. Reducing the number increases the recognition speed.
   - To specify the bar code minimum length in characters, modify the value in the **Min. number of chars** field.
   - To specify the bar code maximum length in characters, modify the value in the **Max. number of chars** field.

- To accept bar codes only if all characters are recognized, that means no rejects select **Return only decodable barcodes**.

4. On the **Restrictions** tab, complete the following substeps as required.
   - To recognize bar codes, select **Barcode** and then select the bar code types you expect to find.

     Note: Do not check more options than required as this slows down processing and affects accuracy.

   - To recognize patch codes, select **Patchcode**.

     Note: 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.

   - If all bar codes or patch codes are oriented horizontally, select **horizontal**.
   - If all bar codes or patch codes are oriented vertically, select **vertical**.
   - If the barcodes can be aligned both horizontally and vertically, select **both**.
   - To detect white bar codes on a black background, select **white**.
   - To detect black bar codes on a white background, select **black**.
   - Too detect both white bar codes on a black background and black bar codes on a white background, select **both**.
5. On the **Preprocessing** tab, complete the following substeps as required.
   - To automatically remove small speckles from the reading zone, select **No isolated pixels**.
   - To automatically fill small gaps in horizontal lines, select **No hor. single pixel gaps**.

     Note: This option is recommended to process bar codes with a corresponding orientation.

   - To automatically fill small gaps in vertical lines, select **No ver. single pixel gaps**.

     Note: This option is recommend processing bar codes with a corresponding orientation.

   - 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 either **horizontal**, **vertical**, or **isotropic**, which means all directions.
   - 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 either **horizontal**, **vertical**, or **isotropic**, which means all directions.
6. Click **Apply** and then click **OK**.

About the OCR Field Level Digit Booster Engine

The digit booster engine 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. The engine supports multi-page documents, but the performance may decrease for multi-page documents with many fields and candidates. Resolutions higher than 300 DPI or lower than 96 DPI lead to boost failures.

The Format Analysis Engine delivers candidates for the Brainware Extraction Engine, so that re-learning is not required after activating the digit booster engine. For digit to digit boosting re-learning is definitely not useful. For alpha to digit boosting relearning might improve the results.

Note: The Digit booster engine requires a [Field Digit Booster] entry in the license file.

About 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).

Activate the Digit Booster Engine

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

1. Switch to **Definition Mode**.
2. In the left pane, on the **Classes** tab, select a class.
3. In the left pane, on the **Fields** tab, select a field.
4. In the right pane, on the **Field** tab, select **Boost field**.
5. Optional. Select **Boost digits only**.

Digit Booster Engine Options

The digit booster engine provides the following options.

Boost field

Enables the alpha and digit to digit boosting to correct a wrong recognition of numbers as letters. Use this option for numeric fields such as amounts or PO numbers only.

Examples

- 0: OCRed as letter 'O'
- 5: OCRed as 's'
- 3: OCRed as 8

Boost digits only

Enables digit to digit boosting to correct a wrong recognition of numbers as numbers, such as 3 OCRed as 8. Use this option for alpha-numeric fields such as invoice number or date.

Enable Extra Boosting

By default, the boost process takes place in the `Document_PostExtract` event. If extra boosting is needed between the analysis and evaluation processes of the extraction, create the registry key ExtraBoostingPriorToEvaluationEnabled. Complete the following steps.

1. Launch the Windows registry editor.
2. Complete one of the following substeps.
    1. For a 32-bit machine, navigate to *[HKEY_LOCAL_MACHINE\SOFTWARE\Oracle\Cedar]*.
    2. 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, enter `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, enter `1`.

About Boosting Tables

The Boost digits only option does not work for table fields. This limitation can be circumvent using a script. For table fields, use the `pTable.IsColumnNumeric` column-level property at the very end of the `Document_PostExtract` event. Set the `isColumnNumeric` property to False to disable alpha to digit boosting.

Example: Boosting tables

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.

About Data Extraction

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

- Create the 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 certain predefined threshold. You can influence the threshold values and the evaluation algorithms, but in most cases, the default settings should be fine.

For now, the following background knowledge is sufficient.

- 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 so-called 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, see About project-level default classes.

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.

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.

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.

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.

Table fields

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

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, enter 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.

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, see About the Brainware Table Extraction engine.

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, complete the following steps.

The class inherits the display name to all its subclasses.

1. On the toolbar, click the **Switch to Definition Mode** button.
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, enter the display name for the field.
4. From the **File** menu, select **Save Project**.

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.

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 then click the table icon.

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, enter a new name and then click **OK**.
4. If required, create a new learnset for the field and repeat the training.

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.

About 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.

The project-level recognition settings affect the following steps.

- The classification step
- Extraction steps using the Format Analysis engine
- Extraction steps using the Zone Analysis Engine, provided that no different settings are specified at the zone level.

Project-level settings affect every document processed with this project. They influence both quality and performance. Therefore, be very careful 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 at the project-level to change the default engines for each recognition type.

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.
    1. On the toolbar, click the **Classify / analyze current document** ❗ button.
    2. Click any of the highlighting buttons and carefully observe the results of your changes.
    3. Repeat the previous steps with a couple of sample documents.
7. If learning was processed before, repeat the training.

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**.

   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, see About the OCR field level digit booster engine.
8. To check the results of field boosting, see About the field-level digit booster candidate evaluation.

Specify the 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, see About 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.
    1. In the **Selected Methods** list, select **Box and Comb Removal**.
    2. On the **Box or Comb Removal** tab, select both **Box Removal** and **Automatic**.
6. If you selected **Lines Manager**, complete the following substeps.
    1. In the **Selected Methods** list, select **Lines Manager**.
    2. 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 an 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.

About 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.

About Line Extraction

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.

Switch Between the Line Extraction Methods

To switch between the line extraction methods, complete the following steps.

1. From the **Options** menu, select **Settings**.
2. In the **Settings** dialog box, on the **Processing** tab, under **Line extraction settings**, select one of the options.

About the 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.

Configure the 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.
5. On the **General** tab, the **Class name format** is used for **Supervised Learning Workflow** (SLW). If you do not use SLW, enter the class name of any one column from the table above in square brackets.

   Example

   ```
   [ID]
   ```
6. In the **Field contents format** field, enter 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.

Example

```
[ID] [Name] [Street]
[City], [State], [ZIP] Tel: [Phone]
```

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.

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.
    1. In the **CSV file** field, enter or browse to the desired CSV file.

       Note: For more information, see About CSV import files.
    2. Optional. If the CSV file is not located in the project directory, clear **Path relative to project file**.
    3. 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.
       If you selected **Import from ODBC data source**, complete the following substeps.
    4. From the **ODBC** source list, select the desired source.
       Note: For more information, see Create an ODBC data source.
        1. In the **User name** field, enter the user name.
        2. In the **Password** field, enter the password.
        3. In the **SQL query** field, enter the query that selects the data to be used in the search pool.
4. 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.

5. Optional. If this directory is not located in the project directory, clear **Path relative to project file**.

6. In the **Pool filename without extension** field, enter 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.

7. Click the **Import** button.
8. On the **General** tab, in the **ID** column, select a unique identifier.
9. If the identifier is alphanumeric, select **Enable alphanumeric identifier**.
10. In the **Search** column, select the fields you want to search for.
11. On the **Import** tab, click **Import**.

About CSV Import Files

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.

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.

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.

   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.

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.

   1. Clear **Enable candidate evaluation using**.
   2. Select **Enable address analysis**.
   3. If the phrase must exist on the document, select the **Phrase is mandatory** check box.
   4. In the table below, in the **Phrase** column, enter the phrase which you expect to find

near the candidate text.

5. 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.

About Multi-Column Attribute Filtering

Both the Brainware V3 core engine 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.

To use the feature with the Verifier search button, configure the .SearchField or the .AnalyzeField call as shown in the following script.

Example: Configure the SearchField or the AnalyzeField

For Associative Search filtering with RTS, configure the extension in the VendorName or VendorASSA field preExtract event. This enables the filtering before the attribute search is implemented in the extraction.

Example: Configure the preExtract event

For more information, see "SCBCdrSupExSettings methods and properties" and in the *Oracle WebCenter Forms Recognition Scripting User's Guide*.

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.

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, enter `ASEnginePoolAllowedCharDifference` and then click **OK**.
5. Right-click the **ASEnginePoolAllowedCharDifference** value and then click **Modify**.

6. In the **Edit DWORD (32-bit) Value** dialog box, in the **Value data** field, to enable the feature, enter zero:`0`.
7. Re-analyze the document.
8. Optional. To disable the feature, delete the registry key **ASEnginePoolAllowedCharDifference**.

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.

   Sample Script

   Partner search

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, see 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

About Learning Lines

The engine considers the following types of 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 “wrong” 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.

About Learning Column Mapping

When learning column mapping, the user trains the engine on how to map the data from the extracted lines to the user's table data. For primary lines, this mapping can diverge for different line classes.

Example

A user learned two different line samples that went to two different line classes internally in one document.

- Mapping rules #1: The user maps “Unit Price” in the document to the “Unit Price” data column and the “Total Price” to the “Total Price” for the first line sample.
- Mapping rules #2: For all lines of the second line type, the user maps “Unit Price” to “Total Price” and “Total Price” to “Unit Price”.

After applying these mapping rules, the BTE engine always uses mapping rules #1 for the lines classified to the first line type and mapping rules #2 for the lines classified as the second line type.

If you have several BTE tables in one WebCenter Forms Recognition class, the engine uses the same learnset for these tables. In other words, if you used interactive learning for one BTE table, the cross-document learning, which happens if the system added the document to the learnset after document validation, applies to all BTE 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.

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**.

Configure Brainware Table Extraction

To set up a table field and define column names for Brainware Table Extraction, complete the following steps. You can create more than one table and configure the Brainware Table Extraction differently for each of them.

1. Switch to **Definition Mode.**
2. In the left pane, on the **Classes** tab, select a parent class.
3. On the **Fields** tab, create a new table field.
4. Right-click the table field and then click **Show Properties**.
5. In the right pane, on the **Analysis** tab, from the **Available Analysis Engines** list**,** select **Brainware Table Extraction**.
6. Complete the following substeps.
    1. On the **Columns** tab, click the **Insert columns** button.
    2. Enter a name for the column.
    3. Optional. If the column must not contain null values, select **Column required**.
    4. Optional. If entries are required for each cell of the column, select **Invalid if empty**.
    5. Optional. If the engine should not convert multiline text to a single line, select **Multiline cells**.
    6. Optional. If you do not want to display the column in Verifier, clear **Visible**.

       Note: The invisible columns must be valid to validate the entire table.
    7. Optional. Repeat the previous steps for all required columns.
    8. Optional. To delete a column, click the **Delete columns** ☒ button.
    9. Optional. To move a column up or down, click the arrow buttons ↑↓.
7. Optional. 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, see Configure Generic Table Extraction.
8. To insert a column format, complete the following substeps.
    1. On the **Labels & Formats** tab, click the **Insert formats** button.
    2. In the **Formats** column, enter the expression. For more information, see What are regular expressions?.
9. From the **Compare Type** list, select a type.

   Note: For more information, see About the Format Analysis Engine.

10. On the **Advanced** tab, in the **Minimal number of words** field, enter the number of words according to the demands of the project documents.
11. Optional. Under **Primary lines extraction**, in the **Minimal threshold** field, modify the value, if for example the correct primary line was not extracted.
12. Optional. Under **Secondary lines extraction**, in the **Minimal threshold** field, modify the value, if for example the correct secondary lines were not extracted.

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 substeps.
    1. 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.
    2. Optional. Select **Derive column definition** to inherit column settings.
    3. Optional. Select **Derive column validation formats** to inherit validation settings.

About Generic Table Extraction

The generic mode of the Brainware Table Extraction performs without prior scripting or training. The extraction bases on the matching of column labels and additionally works with simple mathematical correlation between columns. Furthermore, it increases the speed of the extraction process.

About Column Labels

The Generic Table Extraction identifies the table header line with the help of the column labels. The engine maps the column headers that match with column labels and places matches in the data table under the column name defined in the column list. For multi-line headers, the engine uses only the line judged as the most appropriate for matching.

The search for column labels is a fuzzy search providing some results even for misspelled or poorly OCRed 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.

For example, for the column name Quantity, define the column labels Qty, Quantity, and Quantity shipped.

About Numeric Correlation of Columns

The numeric correlation uses the mathematical interdependency of values to assign the correct values to the corresponding columns. The engine checks the first three columns with the numbers 0, 1 and 2 in the column list on the Column tab for correlation. The correlation follows the scheme 0 * 1 = 2. The columns must be available in the corresponding sequence.

Example for an Invoices project

```
Quantity * Single Price = Total Price
```

Column order
0: Quantity
1: Single Price
2: Total Price

Note: Use this feature only when you expect the columns to correlate in most documents.

Configure Generic Table Extraction

To configure the Generic Table Extraction, 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 substeps.
    1. To insert a column label and activate the generic table extraction, click the **Insert labels** ⬚ button.
    2. Type a unique name for the label with at least three characters.
    3. Repeat the previous steps for all required column labels.
3. Optional. To enable correlation, complete the following substeps.

    Note: For more information, see [About numeric correlation of columns]().

    1. On the **Advanced** tab, select the **Correlation (enable/disable)** check box.
    2. 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.

Limitations

- The user cannot control the identification of a label line (table header line) nor 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, e.g. "line" + "number", and one single expression "line number".

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

- Designer Verifier Test Mode
- Designer Verifier Train Mode
- Verifier
- The column mapping control in context of Brainware Table Extraction interactive learning
- The column mapping and swapping operations in Verifier

Specify Display Names for Columns

By default, the original name displays for columns. To set up customized display names for table column names used in Verifier, complete the following steps.

1. Switch to **Definition Mode.**
2. In the left pane, on the **Classes** tab, select a class.
3. Switch to **Verifier Design Mode**.
4. In the middle pane, right-click the table and then click **Properties**.
5. In the **Properties** dialog box, click **Column Settings**.
6. In the **Table Columns Properties** dialog box, complete the following substeps.
    1. In the **Display Name** column, enter the display name.
    2. If required, repeat the previous steps for other columns and then click **OK**.
7. In the **Properties** dialog box, click **OK**.

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.

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, see Field types.
5. On the **Check Analysis** tab, modify the values for **Threshold** and **Distance**. For more information, see About threshold and distance.

Field Types

The Check Analysis Engine extracts the following fields types from a check document.

MICR (Magnetic Ink Character Recognition) Line

Extracts the bank, account and check number information line from the bottom of the check.

Courtesy Amount Recognition (CAR)

Returns the numeric check amount as a dollar amount.

Legal Amount Recognition (LAR)

Returns the text check amount as a dollar amount.

Cross-Checked Amount

The check amount as a dollar amount, determined using the CAR and LAR results.

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, see “PCDateHint” in the *Oracle WebCenter Forms Recognition Scripting User's Guide*.

Check Number

The check number, returning the position which is generally in the upper right.

Remitter Location

If the engine detects a candidate, the field value is `[Location found]`. For more information, see 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.

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.

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, see 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, see “PayeeVocCoverage” in the *Oracle WebCenter Forms Recognition Scripting User's Guide*.

The combination of the parameters has considerable effect on Payee recognition accuracy.

For more information on configuration of vocabulary parameters, see “PayeeVocEntries” and “PayeeVocCoverage” in the *Oracle WebCenter Forms Recognition Scripting User's Guide*.

You can also define threshold and distance values for the Payee field. For more information, see About threshold and distance.

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.

Payee Vocabulary List

You can define the candidate lists using the table on the Check Analysis tab or by script.

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 it’s weight value overrides the entry created by the user.

If the script adds duplicates, the validation uses the final one added.

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, see “PayeeVocEntries” in the *Oracle WebCenter Forms Recognition Scripting User's Guide*.

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, enter 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** and **Right Arrow**.
4. In the **Weight** column, enter 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**.

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, see “Enable Payee Line script methods” in the *Oracle WebCenter Forms Recognition Scripting User's Guide*.
2. In the script, define the entries using the method `AddItemToVocabulary`. For more information, see “AddItemToVocabulary” in the *Oracle WebCenter Forms Recognition Scripting User's Guide*.
3. Execute the script in your project.

   Note: Candidates added by a script are read-only and display in red.

4. From the **File** menu, select **Save Project** in order to keep the project entries generated by the script.

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**.

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.

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, see About the rules for string construction from words.

The engine provides the following methods to find strings.

String compare

String Compare is a very simple search method that finds each literal occurrence of the specified format string.

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

0 errors	1 error	2 errors
invoice	invoike	invoke
involce	involve	

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

1st trigram	2nd trigram	3rd trigram
Brainain	bra	rai
Trainain	tra	rai

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 one digit. Example: ### matches 123.
@	Represents one upper-case or lower-case letter. Examples: @@@@@@@ matches Capture and @# matches U2.
?	Represents one alphanumeric character. Example: ?rain matches brain and train.
‘	Indicates a word start or end. Example: ‘ABC’ matches ABC, but not ABCD
[ ]	Indicates a number or a range of repetitions of the previous character. Examples: #[3] matches any three numbers like 123. a[1,2] matches a and aa. me[1,2]t matches met and meet.

Regular expressions

Use regular expressions to precisely specify complex format patterns. Designer supports a subset of regular expressions. For more information, see What are regular expressions.

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.

About 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.

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. 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 **Analysis** tab, from the **Available Analysis Engines** list**,** select **Format Analysis Engine**.
5. On the **Format Analysis** tab, complete the following substeps.
    1. Click in the **Format Strings** table and enter the search string.
    2. From the **Compare Method** list, select a method. For more information, see Compare methods.
    3. Optional. Select **Find Designator** and from the **Designator Type** list, select a type. For more information, see Designator types.
    4. In the **Prefix** field, enter the characters or symbols you want the engine to ignore when found at the beginning of the word.

       Example If "$" is entered, $22,99 matches if you search for 22,99.
    5. In the **Suffix** field, enter the characters or symbols you want the engine to ignore when found at the end of the word.
    6. In the **Ignore Characters** field, enter the characters you want the engine to ignore when found anywhere within the word.

       Example If "-" is entered, 2134789 matches if you search for 213-4789.
    7. To define additional strings for the current field, repeat the previous steps. You can combine search methods freely. However, a word can only be member of one candidate.
    8. Optional. To change the order of the format strings, drag the row number to a new position. The engine evaluates the format strings in the order they

appear in the table.

9. Optional. To delete a string, click the row number and then press **Delete**.

6. Initially, most of the controls in this tab are disabled. You need to specify the format string before you can adjust the search method.
7. In the **Format Strings** table, the last row is always empty. To enter a new search string, click in this row number and enter a search string.

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.

About the Rules for String Construction from Words

The Format Analysis Engine may incorrectly fragment an OCRed candidate into several words, or concatenate several words. Therefore the string search includes combinations and subsets of words. The corresponding rules apply to all format strings assigned to a given field.

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, see Rules for string construction from words.

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 computes as follows. `Compare Distance =`

```
 1.00 - Number of Characters (Search String) / Number of
                        Characters (Word)
```

A match requires that the actual compare distance is less or equal the maximum

compare distance. Example

Search String: 1234 (4 characters) Max. Compare Distance: 0.30 Word 1: $1234 (5

characters)

Compare Distance 1: 1.00 - 4/5 = 1.00 - 0.80 = 0.20 < 0.30 -> match

Word 2: $1234.00 (8 characters)

Compare Distance 2: 1.00 - 4/8 = 1.00 - 0.50 = 0.50 > 0.30 -> no match

Max. Wordcount

Allows you to extract strings that

consist of multiple words. Example

By default, the hyphen 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 three or more is required.

Max. gap between words

Specifies the maximum distance in mm that permits word concatenation during the search. Note that the requirements strongly depend on font size.

Max. Candidate Length

Specifies the maximum length in mm a candidate is allowed. 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 enter the spaces in your format string.

Format String	Keep spaces between connected words	Found phrase	Extracted
"Phrase1Phrase2"	Off	"Phrase1" "Phrase2"	"Phrase1Phrase2"
"Phrase1Phrase2"	On	"Phrase1" "Phrase2"	No
"Phrase1 Phrase2"	Off	"Phrase1" "Phrase2"	No
"Phrase1 Phrase2"	Off	"Phrase1Phrase2"	No
"Phrase1 Phrase2"	On	"Phrase1" "Phrase2"	"Phrase1 Phrase2"
"Phrase1 Phrase2"	On	"Phrase1Phrase2"	No

Restrict the Analysis to Specific Areas

To restrict the analysis to specific areas within the documents, complete the following steps.

Note: This may increase processing speed and accuracy.

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. In the right pane, on the **Analysis** tab, select the **Regions** tab.
5. Select **Restrict engine to region**.
6. Select the required pages.
7. For each selected page, specify the percentage you want to take into account.

About the 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 zone analysis 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 format analysis 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
- Cell Limitations
- Reading tables with interleaved columns is not supported.
- Reading tables with empty rows is only supported if the table has a footer line.

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, see “RowNumber” in the *Oracle WebCenter Forms Recognition Scripting User's Guide*.

About the Settings Table

To configure the table analysis, you need to define a settings table. The settings table contains 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.

1. Identifies the document table that best matches the settings table.
2. Analyzes the rows and columns of the document table.
3. Maps the columns of the document table to those of the settings table.
4. Extracts the content of the document table.

Each part of the analysis can create its own set of candidates. In the user interface, the engine returns only the table candidate with maximum confidence so that learning is not required. However, the workdoc contains a list of all table candidates, sorted by confidence. You can access those using 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.

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 substeps.
    1. Click the **Insert columns** button.
    2. In the **Name** column, enter a name for the column.
    3. Optional. If the column must not contain null values, select the **Column Required** check box.
    4. Optional. If entries are required for each cell of the column, select the **Entry Required** check box.
    5. Optional. If the engine should not convert multiline text to a single line, select the **Multiline Cells** check box.
    6. 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, see “ColumnVisible” in the *Oracle WebCenter Forms Recognition Scripting User's Guide*.

7. Optional. To delete a column, click the **Delete columns** ☒ button.
8. Repeat the previous steps for all required columns.
9. 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.

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.

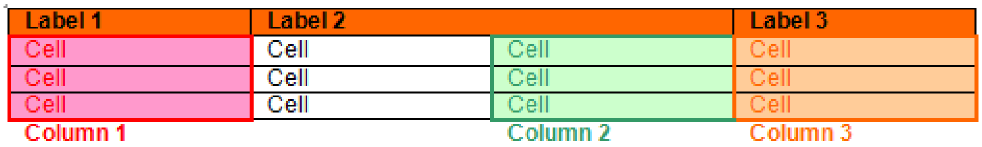
In most cases, all column labels are positioned at about the same horizontal position that means they belong to one text line. This line is called label line. In some cases, the labels extend over more than one text line.
The engine can process single-line as well as multi-line labels.

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.

Example
In the following table, the engine would not recognize the third column.



Note: Since the label detection is vital for table recognition, the OCR quality is very important to obtain good results. Normally, light labels on a dark background are more difficult to read than vice versa.

Define Column Labels and Formats

To define column labels and formats, complete the following steps.

For multi-line labels, no special settings are required as they are automatically combined to one label during table analysis.

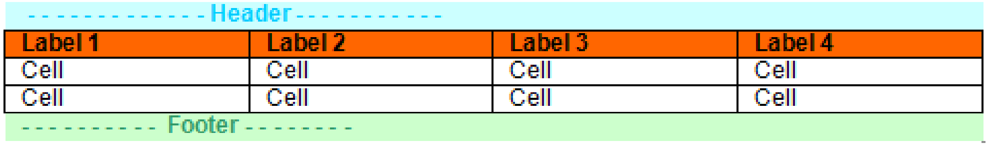
1. In the right pane, on the **Labels & Formats** tab, from the **Column Name** list, select a column.
2. Click the **Insert label** button.
3. In the **Label** column, enter a name for the label.
4. Repeat the previous steps for all required labels.
5. Optional. To delete a column, click the **Delete columns** button.
6. Optional. To specify column formats, complete the following substeps.
    1. Click the **Insertformat** button.
    2. In the **Formats** column, enter the format string.
    3. In the **Compare Type** column, select a type. For more information, see Compare methods and Regular expressions.
    4. Repeat the previous steps for all required format strings.

About Header and Footer Lines

If a header line is available above the label line, the engine can use it to detect the table start position alternatively or in addition to the label line.

Footer lines are located below a table. An example of a footer line is a row that contains the sum of all column cell entries. If a footer line is available, the engine can use it to detect the end of a table instead of or in addition to further criteria.

If your tables contain empty rows, you need a footer. Otherwise, there is no way to determine the table bottom.



- - - - - - - - - - - - - Header - - - - - - - - - - -

Label 1	Label 2	Label 3	Label 4
Cell	Cell	Cell	Cell
Cell	Cell	Cell	Cell

- - - - - - - - - - Footer - - - - - - - -

Define Header and Footer Lines

To define table headers and footers, complete the following steps.

1. In the right pane, on the **Header & Footer** tab, click the **Insert header** button.
2. In the **Header** column, enter a name for the header.
3. Optional. To delete a header, click the **Delete header** button.
4. Click the **Insert footer** button.
5. In the **Footer** column, enter a name for the footer.
6. Optional. To delete a footer, click the **Delete footer** button.

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.

Define the Column Layout

To define the column layout, complete the following step.

1. On the **Analysis** tab, on the **Column Layout** tab, complete one of the following actions.

Situation	Steps
The column position is variable.	1. Select the **Variable column position** option. 2. Optional. 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.
The column position is fixed.	1. Select the **Fixed column position** option. 2. 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.

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, enter a number to skip a number of text lines below the tabletop.

   Note: Use this feature only for tables with a fixed layout.

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, see About header and footer lines.
- Searching for empty cells in a control column. You can also combine these methods.

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 percent.
  - Select Textline distance larger and enter 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.

About Graphical Lines

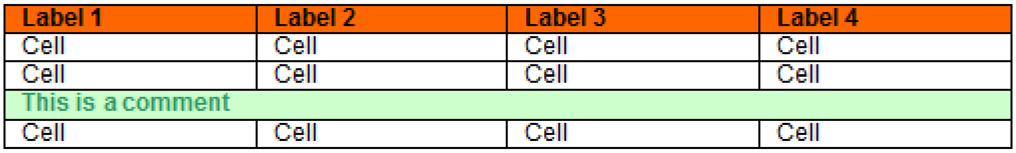
In some cases, tables contain vertical or horizontal graphical lines. Horizontal lines can be used to detect the end of a table instead of or in addition to further criteria. Vertical lines can indicate the segmentation of columns, but are currently not evaluated.

Comment Lines

Comment lines are automatically ignored if the following 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 interfere with the table analysis.

Example



Label 1	Label 2	Label 3	Label 4
Cell	Cell	Cell	Cell
Cell	Cell	Cell	Cell
This is a comment			
Cell	Cell	Cell	Cell

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.

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 SLW process, it is necessary to activate the Brainware Table Extraction engine for the 'Table' field at the child class level.

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 substeps.
   1. 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.

   2. Optional. Select **Derive column definition** to inherit column settings.
   3. Optional. Select **Derive column validation formats** to inherit validation settings.

About Column Mapping

After you select the columns from which you want to extract information for the workdoc in the column 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, see Determine the top of the table.

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.

About Table Mapping in a Script

For information about table mapping using scripts, see “SCBCdrTable” in the *Oracle WebCenter Forms Recognition Scripting User's Guide*.

Zone Analysis Engine

Zone analysis is suitable for extracting data located at fixed positions within the documents, such as with classical forms.

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, bar code 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.

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 reading zone types 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 bar code recognition. The number or strings represented by the bar code 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.

Create a 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 substeps.
    - To create an OCR reading zone, click the **Insert OCR zone** ab button.
    - To create an OMR reading zone, click the **Insert OMR zone** ☒ button.
    - To create a bar code 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.

Adjust a 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**.

About the 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.

Modify the 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

Modify the Settings on the 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 substeps.
    - In the **Name** field, enter 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.

Modify the 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 substeps.
    1. In the **Selected Methods** list, select **Box and Comb Removal**.
    2. On the **Box or Comb Removal** tab, modify the settings.
6. If you selected **Lines Manager**, complete the following substeps.
    1. In the **Selected Methods** list, select **Lines Manager**.
    2. Optional. On the **General** tab, to remove horizontal lines, select **Remove hor. lines** and then enter a minimum line length in millimeters in the **line - horizontal** field.
    3. Optional. To remove vertical lines, select **Remove ver. lines**. Enter a minimum line length in millimeters in the **line - vertical** field.
    4. Optional. Select **Repair Characters**.
7. To test the effect of the methods, complete the following substeps.
    1. Ensure that the current document is visible and not hidden behind the dialog box.
    2. 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.
    3. View the document to check the preprocessing effects.
    4. Optional. If you want to repeat the test, click the **Reset** button before you proceed.

Modify the 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.

Test the Zone

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.

About Anchors

It is quite common for the alignment of scanned paper documents to vary. While OCR can correct for rotation, the horizontal and vertical position will often fluctuate from page to page.

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.

Create an Anchor

To create an anchor, complete the following steps.

1. Load a document that belongs to the respective class.
2. Switch to **Definition Mode.**
3. In the middle pane, click the **Insert anchor** ⚓ button.
4. On the document, drag to mark the anchor zone.

   Note: Ensure that the anchor zone is large enough to contain the geometric feature, while not overly large. The anchor position should be close to the geometric feature you want to use.

5. Optional. To create more anchors, repeat the previous steps.

Modify the 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 substeps.
   - In the **Name** field, modify the anchor name.
   - 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.
3. To specify for which geometric feature the anchor is looking for, from the **Anchor Type** list, select one of the following types and then modify the properties of the anchor type, if applicable.
   - **Fit to image corner**.

     You can position the anchor anywhere on the document page.

   - **Fit to Text**.
   - **Fit to line**.
   - **Fit to box corner**.
   - **Fit to "T"**.
   - **Fit to cross**.
4. Click the **Learn** button.

   A message will indicate learning was successful and the learned geometric feature will be indicated by dotted green lines. If unsuccessful, modify the position or size of the anchor and then click **Learn** again.

5. Click **OK**.

Note: Any change to the anchor position or size removes learning.

Configure the Anchor Type Fit To Text

The anchor type **Fit to Text** allows to locate a specific document search text. To configure this anchor type, complete the following steps.

1. Position and size the anchor so that it contains only the search text.
2. Right-click the anchor and then click **Properties**.
3. In the **Anchor Properties** dialog box, click **Text Recognition Settings**.
4. In the **Text-Anchor OCR Properties** dialog box, configure the required recognition engine and then click **OK**.
5. In the **Anchor Properties** dialog box, click **Learn**.

   The learned search text appears within Properties and is outlined by a dotted green box on the document image. The anchor is then be enlarged automatically to increase the likelihood of locating the search text on subsequent documents.
6. Optional. Further enlarge the anchor by resizing the anchor on the document image.
7. If you later need to relearn the anchor, repeat the previous steps, so that the anchor searches for only the search text and not all text within the enlarged anchor.

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.

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 substeps.
   1. 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.
   2. 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.

About Validation

The validation and output formatting features enable you to standardize the development and maintenance cycles for each WebCenter Forms Recognition implementation. However, standard validation has not replaced scripted validation. WebCenter Forms Recognition still supports validation scripts. For a brief introduction to script validation, see About validation scripts. For more information, see 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 processes 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 is 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 only performs when the standard validation is valid. If one document of a batch is invalid, the system passes the whole batch to Verifier. If a field’s standard validation is set to invalid, the corresponding script events will not execute.

About the 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.

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.

About the 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.

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.

Modify the 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**.

About 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.

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
- Checkbox
- List
- Table

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*.

About Validation Rules for Classes

Validation rules for classes provide the following possibilities.

- Override or accept the parent-level settings for the “Force Validation” option.
- Accept all validation settings derived from a parent or from the child class.
- Establish general validation rules for the class.

Modify the Validation Settings for a Class

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, see 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, see About derived validation.
8. Optional. On the **General** tab, to build an expression, complete the following substeps.
    1. To copy the field name to the expression, double-click an amount field.
    2. Click one of the function buttons. For more information, see Functions for validation rules.
    3. Click one of the operator buttons. For more information, see Operators for validation rules.
    4. If required, enter the parameters.
    5. To add the expression to the list of validation rules, click **Add**.
    6. 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**.

Modify the 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 substeps.
    1. From the **Validation Rules** list, select a validation rule.
    2. Modify the rule.
    3. Click the **Update** button.

Operators for Validation Rules

Operator	Description
=	Equal
>	Greater than
<	Less than

*	Multiplication
/	Division
+	Plus
-	Minus
(	Open grouping for expression
)	Close grouping for expression
!	NOT
!=	Not equal
&&	AND
||	OR

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	Absolute value of expression

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`, always use a period as decimal separator and do not use thousand separators.

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

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.

Modify the 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, see 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 substeps.
   - 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.

Modify the General Validation Rules for a Text Field

To modify the general validation rules for a text field, complete the following step.

- In the right pane, on the General tab, complete any of the following substeps.
  - 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 enter the length in the field below.
  - To define a maximum length for the field, select Max. Len and then enter the length in the field below.

Note: For more information about the settings, see General validation settings for a text field.

Modify the 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, see General validation settings for an amount field.
2. From the **Region** list, select a region.
3. Optional. Complete the following substeps.
   1. Select the required settings from the lists in the right column.
   2. To copy all settings selected in the right column to the left column, click the **<< All <<** button.
   3. Optional. In the **Name** field left to the **<< All <<** button, enter a name for the rule.
4. Optional. Complete the following substeps.
   1. Select an entry from one of the lists in the right column and then click **<<**.
   2. 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”.
   3. 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.
   4. Optional. In the **Name** field left to the **<< All <<** button, enter a name for the rule.
5. Optional. Enter 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.

Modify the 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 substeps.
    - 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**.

Modify the 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 **Checkbox Caption** field, enter a free text check box caption, such as Yes/No, Accept/Decline, or True/False.
2. In the **Checked Value** field, enter the value for a selected check box according to the value you specified in the **Checkbox Caption** field.
3. In the **Unchecked Value** field, enter the value for a cleared check box according to the value you specified in the **Checkbox Caption** field.
4. Complete one of the following substeps.
    - To specify whether a check box is selected by default, select **Checked**.
    - To specify whether a check box is cleared by default, select **UnChecked**.

Modify the 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 enter the length in the field below.
2. Optional. To define a maximum length for the field, select **Max. Len** and then enter the length in the field below.
3. To add a value to the list, click the **Insert** button and then enter 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**.
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**.

Modify the 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 substeps.
    1. To specify the type the engine uses to validate the table field, in the **Type** column, select an entry from the list.
    2. Optional. To assign a template to the field, in the **Template** column, select an entry from the list.

About Character Filtering

The Char Filtering tab is available for all types of fields. This tab enables you to specify which characters are valid for the field, which characters you want to remove and which you want to replace during validation. This tab also enables you to specify the OCR confidence and case conversion.

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

Modify the Character Filtering Settings

To modify the character filtering settings, complete the following steps.

1. Under **Valid Characters**, complete one of the following substeps.
   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, enter 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 substeps.
    1. In the **Old Char** column, enter the character you want to replace.
    2. In the **New Char** column, enter the character you want to insert.
4. Optional. Complete the following substeps.
    1. To accept all OCR results, regardless of quality, select **No Rejects**.
    2. 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**.

General Validation Settings for a Text Field

Multi-Line

Converts multi-line data to a single, non-delimited line.

Allow Empty Field

The field text can be empty.

Min. Len

Defines the minimum length of the field.

Max. Len

Defines the maximum length of the field.

General Validation Settings for an Amount Field

Amount Range

Select this setting to enable the Min. Amount and Max. Amount fields for validating a range of values.

Min. Amount

Fields with lower values are invalid.

Max. Amount

Fields with higher values are invalid.

Region

Provides predefined regional amount settings.

Currency Symbol

Currency symbol for the selected region.

Positive Format

Uses the currency settings for the selected region and standard currency formats for depicting positive numbers.

Negative Format

Uses the currency settings for the selected region and standard currency formats for depicting negative numbers.

Decimal Symbol

Decimal separator. This options works only in combination with “Decimal Places”.

Note: You cannot use the same symbol as for the thousands separator.

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”.

1000 Separator

Thousands separator. This options works only in combination with “Digit grouping”.

Note: You cannot use the same symbol as for the decimal symbol.

Digit grouping

You can select to have two or three characters per group or not to group digits at all. This options works only in combination with “1000 separator”.

General Validation Settings for a Check Box Field

Checkbox Caption

Free text check box caption, such as Yes/No, Accept/ Decline, or True/False.

Checked Value

Specifies the value for a selected check box.

Note: Ensure that this setting does not conflict with the “Checkbox Caption”.

Unchecked Value

Specifies the value for a cleared check box.

Note: Ensure that this setting does not conflict with the “Checkbox Caption”.

Default (Checked/ Unchecked)

Specifies whether a check box is selected or cleared by default.

General Validation Settings for a List Field

Min Len

Defines the minimum valid field length.

Max Len

Defines the maximum valid field length.

Insert/delete list items

Create or remove list items.

Allow list values only

Only values that appear on the list are validated.

Fill list of Suppliers

This option is only available for fields which are designed to use Associative Search Engine (ASE). Populates the list with the Associative Search results.

Show Nearest Value

Automatically selects the most matching entry from the list while typing in Verifier or Designer's Verifier Test Mode.

Sort list items

Sorts the items in alphanumeric order.

General Validation Settings for Table Fields

Type

Indicates whether the engine validates the table column as text, amount, check box, list, or date.

Template

Assigns a template to the field.

About Output Formatting

Output formatting enables you to standardize the format for date and amount fields.

You must specifically enable the output formatting at the field level. The formatted values are saved as the `FormattedText` for the specified field and can be accessed in script.

For more information, see Validation > Validation Rules for Fields > Validation Scripts.

Modify the Output Formatting Settings for a Date or an Amount Field

To modify the output formatting, complete the following steps.

For more information, see Output Format validation settings for a date field and 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**.

Output Format Validation Settings for an Amount Field **Region**

Provides country-specific predefined amount settings.

Currency Symbol

Currency symbol for the selected region.

Positive Format

Uses the currency settings for the selected region and standard currency formats for depicting positive numbers.

Negative Format

Uses the currency settings for the selected region and standard currency formats for depicting negative numbers.

Decimal Symbol

Decimal separator.

Note: 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.

Note: 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.

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.

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.

Create a 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 substeps.
    1. From the **Validation Type** list, select a type.
    2. Optional. From the **Available Templates** list, select a template and then click **Copy**.
    3. Modify the settings.
5. Click **Save**.
6. In the **New Template** dialog box, enter a name for the template and then click **OK**.

Apply a 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. From the **File** menu, select **Save**.

Create a 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 substeps.
    1. Optional. From the **Available Templates** list, select a template and then click

Copy Template.
2. Modify the settings.
3. Click **Save as Template**.
4. In the **New Template** dialog box, enter a name for the template and then click **OK**.

Apply a 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. From the **File** menu, select **Save**.

Modify a 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. From the **File** menu, select **Save Project**.

Import a 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**.

Export a 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, enter a name and then click **Save**.

Delete a 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**.

Rename a 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, enter a new name for the template and then click **OK**.

Validation Scripts

- About Validation Scripts
- About Output Formatting in Scripts
- Value Property for Simplified Scripting of Extended Validation Routines

About Validation Scripts

WebCenter Forms Recognition offers various script methods and properties to further customize validation. Note that script validation overrides the validation settings made in the Designer GUI. 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, see the *Oracle WebCenter Forms Recognition Scripting User's Guide*.

About Output Formatting in Scripts

You can access the formatted value in the `SCBCdrField` property `FormattedText`:

`pField.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 these properties from script events, such as `ExportDocument` or `Document_Validate`. For more information, see “FormatForExport”, “FormattedText“, “ExportDocument“, and “Validate” in the *Oracle WebCenter Forms Recognition Scripting User's Guide*.

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 a script.

The field or table cell property `Value`, that contains the field value in the variable type depending on the field validation type. The Value property is VARIANT and returns date (for date fields), currency (for amount fields), Boolean (for check box fields), index (long value) 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).`

Below is a script sample that demonstrates usage of the properties in the custom script.

Example: Use the Value property

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.

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).

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.

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, WebCenter Forms Recognition analyzes the areas around each candidate for characteristic properties that can be learned.

Configure the 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.

Brainware Field Extraction Engine for Generic Field Extraction

The Brainware Field Extraction engine introduces a dimension in automatic extraction of header field data on a generic level that means when multiple logical document classes are handled with a single internal knowledge base. The engine uses an advanced N-gram Triton technology and provides high quality field extraction based on the following statistical evaluation.

- 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, which avoids time consuming specification of regular expressions through the Format Analysis Engine.

The engine fully supports the Supervised Learning Workflow.

You can configure and train the Brainware Field Extraction (BFE) engine in the same way as the Brainware Extraction engine. For more information, see Configure the Brainware Extraction engine.

You can assign the engine 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 the BFE settings view also displays the learn status for the selected field on the General tab.

Differences Between the Brainware Field Extraction Engine and the Brainware Extraction Engine

Format Analysis engine preprocessing and time consuming configuration is no longer required for the BFE engine function. See the following limitations and About Brainware Field Extraction candidates in a script for more details.

The Brainware Field Extraction engine does not support microlayout settings.

Limitations

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.

Configure the 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.

About Brainware Field Extraction Candidates in a Script

The following script sample shows how to replace the extracted BFE candidates with the ones extracted by the Format Analysis engine or how to keep both kinds of candidates in the candidates array.

Example: BFE candidates in a script

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 the Supervised Learning Workflow (SLW), due to above mentioned reasons, we recommend to use the Brainware Extraction engine for automatic SLW vendor level training. For this purpose, assign the “Brainware Extraction” engine to the header fields of the SLW’s root classes using a special dedicated class for generic extraction through Brainware Field Extraction engine. For example, the following classes’ hierarchy:

Invoices

- Generic
- VendorClass1
- ...
- VendorClassN

The “Invoices” class is the root template class for SLW training, while the “Generic” class defines generic extraction through BFE engine. In this connection, the standard classification result for “Invoices” has to be set to “Generic”, so that in case a document has not been classified to one of available vendor class with precise layout extraction “VendorClassX”, it goes to “Generic” class with generic extraction pattern defined. After the generic processing has been applied, the classification result has to be set back to “Invoices” in case further SLW processing is to be maintained. This can be achieved by placing the following script code into “PostExtract” event’s handler of the “Generic” document class.

Example: Reset the classification result

In some project configurations, when vendor classification succeeds but does not deliver any results for some selected fields, it may also make sense to take an advantage of the “Retain previous extraction results” option and define double extraction processing to combine generic extraction with vendor level extraction.

About Training Header Fields in Normal Train Mode

Use this feature for quick learning of the 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 SLW to use Verifier Train Mode for the purpose of Brainware Field Extraction learning.

Note that 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 “run- time” 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 internal “Correct Results” Brainware Classifier’s class and using all its other available candidates as multiple members of “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 “Wrong Results” class’ content. This content, in case no candidates were available prior to start of the document learning procedure, will be reduced up 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 / extraction goal in general, it is obviously not optimal in terms of quality.

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.

   Note: The system adapts the area to the best matching position.

6. To create the candidate and apply the required field assignment for further training, double-click the selected area.

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 WebCenter Forms Recognition to learn how to extract the data.

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 a couple of candidates instead of just one 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.

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.

Note: If you use field inheritance, select the parent class.

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.

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.

Create Learnsets for a Single Field

To create learnsets for a single field in a class, complete the following steps.

Prerequisite: Your project settings must make sure that the field selection is not changed when switching to the next document.

1. Switch to **Document Selection Mode**.
2. In the left pane, select the first document from the batch that contains the documents for your learnset.
3. Switch to **Normal Train Mode**.
4. On the **Classes** tab, in the lower left pane, 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 turned on automatically.
5. On the toolbar, click the **Classify / analyze current document** ! button to highlight all candidates for the first field on the document.
6. Complete the following substeps.
   1. In the middle pane, click the correct candidate.

      The candidate's text is written to the field, and the field's background color changes to green.
   2. To add the current document to the extraction learnset, on the toolbar,

      click the **Add document to Learn Set** button.

      Note: The program continues automatically with the next document and the selected field.
   3. Repeat the previous steps until you have enough samples.

Improve the Learnset

To improve the learnset, complete any of the following steps.

- Skip fields.

- To repeat the training of a field, complete the following substeps.
    - In the left pane, on the Fields tab, select a field.
    - On the toolbar, click the Classify /analyze current document button.
    - To add the current document to the extraction learnset, on the toolbar, click the **Add document to Learn Set** 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 substeps.
    - Click the **Highlight Words** button.
    - In the left pane, on the Fields tab, select the required field.
    - In the middle pane, click the word you want to add.
    - In the Add Candidate dialog box, click **Yes**.
    - In the right pane, on the Analysis tab, review the settings to ensure that the candidate can be found.

Modify the Learnset

To modify the extraction learnset, complete the following steps.

1. Switch to **Document Input Selection Mode – Learn Set**.
2. In the upper left pane, right-click a class and then click **View Extraction Learn Set**.
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.

About Learning

During learning, WebCenter Forms Recognition uses your learnsets and the manually assigned candidates to identify the correlation between successful candidates, unwanted candidates and their environment. This correlation teaches WebCenter Forms Recognition to automatically extract data that is similar to what was learned via the candidates.

Learning is required in the following cases.

- After you set up your extraction scheme and created learnsets
- When you added, deleted, moved, or renamed fields
- When you have changed any parameters affecting the analysis and evaluation methods
- When you have changed a learnset

Classification and extraction are usually learned together. To learn extraction, proceed as described for classification. For more information, see Use the learn mode.

Check the Learn Status of a Field

You can check at any time whether a field has already been learned or requires re-learning. To check the learn status of a field, complete the following steps.

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

3. On the **Fields** tab, select a field.
4. In the right pane, on the **Evaluation** tab, on the **Brainware Extraction** tab, check the learn status.

About Testing 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.

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 substeps.
    - 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 substeps.
    1. To highlight all candidates for the currently selected field in maroon, on the toolbar, click the **Highlight Candidates** button.
    2. To view the weight of a candidate, in the middle pane, move the mouse over the candidate.
4. Complete the following substeps.
    1. 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.
    2. 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 substeps.
    - 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.

About Test Results

The test results display in the left pane on the Fields tab. Valid fields display with a white background. Invalid fields display with a red background.

A field is invalid, but not empty, if the best candidate has the required weight, but not the required distance or the best candidate does not meet standard or script-based validation requirements.

A field is invalid and empty if no candidate was found or no candidate has the required weight.

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. For more information, see The Statistic View tab.
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 pane 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.

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.

Extraction Optimization

Most extraction problems are caused by the following problems.

- Incomplete configuration
- Insufficient quality of the learnset
- Insufficient quality of the OCR

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 0, check whether an evaluation method is specified at all. For more information, see Specify the analysis engine.
- Check whether the evaluation has been learned. For more information, see About learning the extraction.
- If you use field inheritance, check the parent and child configuration.

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.

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.

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 WebCenter Forms Recognition. To use the feature, either configure two separate projects or perform the secondary extraction using a script.

For an example using two projects, see Extraction retaining using two projects. For an example using a script, see Extraction retaining using a script.

Extraction Retaining Using Two Projects

Project "P1" contains definition of classes' hierarchy with root "Header Fields" class. "Retain previous extraction results" feature is turned off. A Runtime Server instance applies an extraction workflow step with "P1" redirecting the documents to state 551 in both failure and full success cases. Another Runtime Server instance applies extraction with the second project "P2" that contains the definition of "Table Fields" class. "Retain previous extraction results" feature is turned ON for "P2". This Runtime Server instance has 551 as input extraction state and "normal" 552 / 700 output workflow states for extraction failure / success respectively. The resulting workdocs contain all desired fields "A", "B", "C", and "T".

Extraction Retaining Using a Script

The project "P1" contains both classes' hierarchies: the one with "Header Fields", and the 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 one below.

Example: Extraction retaining

Activate Extraction Retaining

If double or triple document processing is required, activate the extraction retaining. Complete the following steps.

1. From the **Options** menu, select **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**.

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.

Example

	Weight of Candidate A	Weight of Candidate B	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

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.

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.

About Supervised Learning

Use the Supervised Learning Workflow (SLW) to automatically create and learn new document classes. This automatic process creates the new derived document classes for a 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.

You must learn any document you add to the local learnset before it is released. Otherwise, the subclass is created in the project. In addition, save the local project to create the subclass in the global project.

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 Designer 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.

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 **Designer**, open the project.
3. From the **Options** menu, select **Settings**.
4. In the **Settings** dialog box, on the **Verifier Train 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 **Verifier Test Mode**.
9. Test the project-related validation.
10. In **Verifier**, activate the Supervised Learning Workflow.

    Note: For more information, see “Configure Supervised Learning” in the *Oracle WebCenter Forms Recognition Scripting User's Guide*.

Example for a Supervised Learning Workflow

You have an existing vendor class, named “Invoice” that processes invoices from the supplier "Company International Technology, S.A". You have a document class with the same purpose, but with a different name, such as BW1234567, that can be created through the Supervised Learning Workflow, but may cause conflicts.

To avoid conflicts, take the following cases into account to combine non-SLW and SLW processed class nodes.

Criteria	Case A	Case B	Case C	Case D
Existing classes conflict with potential classes	No	Yes	Yes	Yes
Automatic Learning required	NA	No	Yes	Yes
Level 1 class (Derived from Base DocClass)	NA	No	Yes	No
Base class do not satisfy requirements for SLW execution or classification field improperly configured	NA	NA	NA	Yes

Case A

The existing set of classes does not coincide with the set of classes that the Supervised Learning Workflow might create.

This case is the easiest to handle. All you have to do is 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

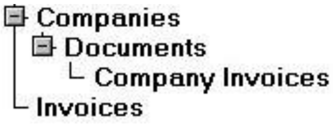
The existing set of classes does coincide with the set of classes that the Supervised Learning Workflow might create. Automatic learning for these classes is not required.

Case B occurs when you want to keep an existing document class with its learnset, scripting, and other settings, 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 would have to switch off automatic learning for this document class.

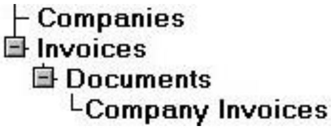
To disable “automatic class creation” for a particular class, set the vendor type field in the CSV file to 2 for the corresponding class search entry. For details, see About the Associative Search Engine.

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 one of the following three values: 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, see the *Oracle WebCenter Forms Recognition Verifier User's Guide*.

You must re-launch the pool import in the Associative Search settings to establish Classification Field column as the Vendor_Type field. Do not use this field as a search field.

In the scenario, the illustration at right shows how a migrated project might look. This illustration shows a sample hierarchy consisting of the old document class Companies with its derived document classes and a new base class conceived for SLW, Invoices.	 Companies Documents Company Invoices Invoices

In the scenario, there is an entry Company in the vendor pool, but learning is prohibited for this entry. Although no new document class for Company will be created, you still must ensure that the system classifies the documents to the "old" document class, Company 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. In the illustration at right, automatic learning for Company Invoices is disabled and all the configurations (such as for 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.	 Companies Invoices Documents Company Invoices
• 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 scripts to ensure that the document is classified to the correct document class	

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.

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 SLW and specified on the ASE settings for Classname 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 Classname Format for the corresponding entry in a CSV file.
- Change Classname 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.

(An example might be My Document Class 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 (For example, vendor name, space, unique vendor ID.) Then, in the settings for the Associative Search Engine, use *[My Document Class Name]* as Classname Format.

The base document classes must satisfy the constraints for execution of the Supervised Learning Workflow: 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, see About Supervised Learning.

Case D

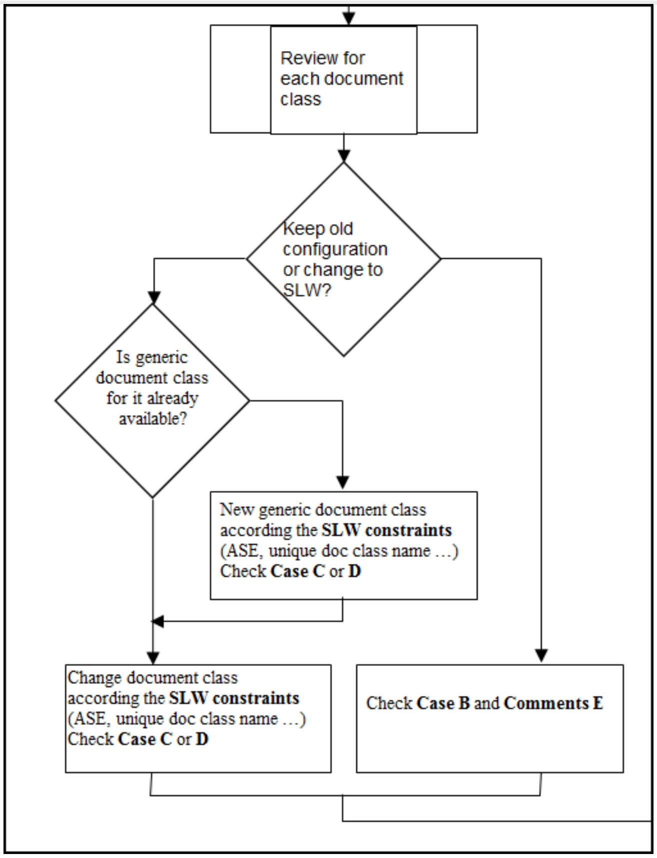
The existing classes do coincide with the set of classes that might be created by the Supervised Learning Workflow. Automatic learning for these classes is required. These classes are not on Level 1 in WebCenter Forms Recognition project classes' hierarchy or their base classes do not satisfy constraints for Supervised Learning – meaning that Classification Field settings are not correctly configured.

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 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 Definition Mode, Classes view, 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.

The following diagram illustrates the steps that the system administrator must consider for each document class when migrating a project.


Review for each document class
Keep old configuration or change to SLW?
Is generic document class for it already available?
New generic document class according the **SLW constraints** (ASE, unique doc class name …) Check **Case C** or **D**
Change document class according the **SLW constraints** (ASE, unique doc class name …) Check **Case C** or **D**
Check **Case B** and **Comments E**

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 set up the classname 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, see About Supervised Learning.

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 substeps.
    - Switch to Definition Mode.
    - In the left pane, on the Classes tab, right-click the base document class and then click Show Properties.
    - In the right pane, on the Classification tab, select an engine.
- To define a class as default classification result, complete the following substeps.
    - Switch to Definition Mode.
    - In the left pane, on the Classes tab, right-click the project and then click Show Properties.

- 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.

- Insert at least one header field (usually called a text field) for which the Associative Search Engine is selected as the analysis engine.
- 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.
- 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.

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, see 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.

Training of Other Fields

For all other text fields that you defined for a generic document class besides the one to which the Associative Search Engine is assigned, you must use the Format Analysis Engine for analysis and the Brainware Extraction Engine for evaluation.

For table fields, use the Brainware Table Extraction engine.

About Creating Derived Document Classes

You do not have to manually create the derived document classes, the Supervised Learning Workflow creates them automatically when it adds the documents to the learnset.

When adding the documents to the learnset, the SLW automatically creates a new derived document class. The new classname corresponds to the settings established for the document-level settings of the field Classname Format on the Analysis Editor. Therefore, this must be defined in a way that ensures uniqueness of every classname that may be created, that means that the ID, or any other unique field, must be part of the Classname Format.

Though not obligatory, it can be useful to train some additional fields for the generic class. that means that extraction results are available for the new class, and the Verification form is therefore already be partially populated.

To train the generic Document Class, 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 you only need to check the extraction results.

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.

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 WebCenter Forms Recognition projects used for invoice processing.
- The 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.

Add Supervised Learning to Brainware Layout Classification

The Brainware Layout Classification engine automatically trains the derived (level 1) document classes. You can also take the results of the Associative Search Engine into account for the classification. 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 **Supervised Learning** tab, select **Make supplier field invalid**.
4. Select one of the following options.
    - **Do smart decision**. WebCenter Forms Recognition determines the right DocClass based on an algorithm that compares the results of the

Associative Search Engine and the Brainware Layout Classification.

- **Assign current document class name to supplier field**. WebCenter Forms Recognition 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 Brainware Layout Classification.
- **Assign supplier field content to current document class name**. WebCenter Forms Recognition uses the result of the Associative Search Engine if the DocClass is available; otherwise it uses the result of the Brainware Layout Classification.
- **Do nothing** WebCenter Forms Recognition uses only the result of the Brainware Layout Classification and not the result of the Associative Search Engine.

5. Optional. To activate the **Correct Tables** button in Verifier, select **Allow interactive table extraction for plain Verifier user**.
6. Optional. To perform local classification and extraction only for invalid fields, complete one of the following substeps.
    - On the toolbar, click the **Classify / analyze current document** button.
    - On the **Supervised Learning** tab, select **In Verifier apply local extraction only for invalid fields**.

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: 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**.

ASE Pool Configuration

The following fields are mandatory

- Vendor ID: A unique numerical designation for the vendor name.
- Internal_ID: A unique sequential designation for each record in the file.
- Vendor_Name: A short form of the Vendor’s full name.
- Vendor: 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 designate another field as vendor type, as long as the values are correct.

Possible values

- 0: Adds all documents classified by that vendor name to the local learnset. This is the default value if no Vendor_Type 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.

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, see "Supervised Learning settings" in the *Oracle WebCenter Forms Recognition Verifier User's Guide*.

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 validate the documents, which are presented as batches. The following rules determine whether a batch is valid.

- 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 needs to correct and confirm all invalid or uncertain results from a given batch.

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, the following 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 and verification is defined by a form. You need to create a set of forms used for verification of data extraction. You can simplify this step by using default forms, but it can also be a complex task if there are special requirements. In addition, you can implement custom forms or verification routines in scripts.

The verification steps take place when a batch of documents reaches a predefined state. 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, see the following documentation.

- *Oracle WebCenter Forms Recognition Runtime Server User's Guide*. The Runtime Server settings determine the input and output states in the WebCenter Forms Recognition workflow.
- *Oracle WebCenter Forms Recognition Verifier User's Guide*. The Verifier settings determine the input and output states of verification steps.

About Verification Projects

Designer organizes verification forms in verification projects. It uses the following structure.

- Each 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 is used by default. The remaining forms are only used when a document that requires validation is in a certain predefined state.

Each verification screen has up to four areas.

- Form area
- Direct input area
- Document viewer area
- User info area

Configure the Project Validation Properties

To configure the project validation properties, complete the following steps.

Prerequisite Classes and fields are already defined.

1. Switch to **Verifier Design Mode**.
2. In the upper left pane, right-click the project and then click **Project Properties**.
3. In the **Validation Settings** dialog box, under **Force Validation**, select one of the following options.
    - **Permitted**. The user can force validation by pressing Enter three times.
    - **Forbidden**. The user cannot force validation.
4. Optional. If you have script code associated with the `ScriptModule_RouteDocument` event, modify the states to test your scripting in Verifier Test Mode.

   Note: These states are not used in Verifier.

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 is visible in Verifier.
- You can create forms for all classes in the project in one step.

- You can change fonts, colors, and background colors for verification forms and elements. For more information, see “SCBCdrVerificationForm“ in the *Oracle WebCenter Forms Recognition Scripting User's Guide*.

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

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.

Each control must be assigned to a data extraction field. The viewers display the document area that Designer extracted to fill a given field.

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.

Verifier Design Mode Toolbars

The Verifier Design Mode toolbars on the top and the right side provide various functions to work with verification forms.

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. • 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. • 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. • 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. - 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.
A	Add New Label Element
☐	Add New Viewer Element
OK	Add New Button Element
✗	Delete Selected Elements
🔑	Define Smart Indexing
🔑	Delete Smart Index
⚡	Show Action List
🔒	Lock Verifier controls

Create an Empty Form

To create a single empty form, complete the following steps.

1. Switch to **Verifier Design Mode**.
2. In the upper left pane, right-click a class and then click **Insert Empty Form**.

   The form displays in the middle left pane and the fields of the class display in the lower left pane.
3. The form background provides a grid to simplify aligning form elements. Use the **Enlarge grid step** and **Shrink grip step** buttons to adjust the size.
4. Optional. Click the **Show/hide grid** button to turn the grid on or off.
5. From the lower left pane, drag the required fields to the form.
6. Add a label to each field.
7. Optional. Add other form elements.
8. Optional. To move a group of fields to another place on the form together, drag a rectangle around the fields, press Shift, and then drag the group.
9. Optional. To prevent moving elements accidentally, click the **Lock Verifier Controls** 🔒 button.
10. Optional. Configure the form validation properties.

Create a Default Form for a Class

You can create a form that contains all fields of the class as well as the corresponding labels and viewers. To create a default form for a class, complete the following steps.

1. Switch to **Verifier Design Mode**.
2. In the upper left pane, right-click a class and then click **Insert Default Form**.

   The form displays in the middle left pane and the fields of the class display in the lower left pane.
3. The form background provides a grid to simplify aligning form elements. Use the **Enlarge grid step** and **Shrink grip step** buttons to adjust the size.
4. Optional. Click the **Show/hide grid** button to turn the grid on or off.
5. Optional. Add other form elements.
6. Optional. To move a group of fields to another place on the form together, drag a rectangle around the fields, press Shift, and then drag the group.
7. Optional. To prevent moving elements accidentally, click the **Lock Verifier Controls** button.
8. Optional. Configure the form validation properties.

Create a Default Form for Each Class in the Project

To create a default form for each class, complete the following steps. Note that this deletes all previously defined verification forms.

1. In the upper left pane, select the project and then click the **Insert New Default Verification Forms for all DocClasses** button.
2. In the **Insert Default Verification Forms** dialog box, click **Yes**.
3. In the upper left pane, select a class and review the created form.
4. The form background provides a grid to simplify aligning form elements. Use the **Enlarge grid step** and **Shrink grip step** buttons to adjust the size.
5. Optional. Click the **Show/hide grid** button to turn the grid on or off.
6. Optional. Add other form elements.
7. Optional. To move a group of fields to another place on the form together, drag a rectangle around the fields, press Shift, and then drag the group.
8. Optional. To prevent moving elements accidentally, click the **Lock Verifier Controls** button.
9. Optional. Configure the form validation properties.

Configure the Form Validation Properties

To configure the form validation properties, complete the following steps.

1. Switch to **Verifier Design Mode**.
2. Display a form and ensure that no form elements are selected.
3. In the forms pane, right-click on the background and then click **Properties**.
4. In the **Properties** dialog box, complete any of the following substeps.

- To use this form only for documents with a given state, from the **Process Step Filter** list, select a state or type a value. For default forms, keep the initial value of -1.
- To use the form as the default form, such as in all verification steps without a special predefined input state, select **Default Form**.
- If only one user verifies an entire batch, select **Complete validation required**. 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 several users validate a batch in stages, select **Partial Validation allowed**. If partial validation is allowed, the user must validate only the fields on the form. With partial validation, you can employ multiple forms, that means one for each stage. You can set an output state for partially validated documents in Designer, which overrides the output state set in Verifier.

  Note: The **Name** field, which displays the name of the form, and the **Assigned DocClass** field, which displays the name of the associated document class, are read-only.

5. Click the **Validation Settings** button.
6. In the **Validation Settings** dialog box, modify the settings if required.

   Note: For more information, see Modify the validation settings for a class.

Display a Form

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.

Display all Forms of the Project or a Class

To display all forms of the project or a class, complete the following step.

- Complete one of the following actions.

Situation	**Steps**
Display all forms of the project.	- From the Edit menu, select Verification Form > Show All.
Display all forms of a class	- In the upper left pane, select a class.

The forms of the project or the selected class display in the middle left pane and provide the following information.

Column	Description
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. 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 Default if no specific input state has been selected.
Out Filter	Displays the output state assigned to a form, or Default if no specific output state was selected. Output states are relevant if partial validation of documents is allowed and are for example useful in the context of exception handling.
Assigned to	Displays the name of the assigned document class.

Delete a Form

To delete a single form, complete the following step.

- In the middle left pane, right-click a form and then click Delete Form.

Delete all Forms

To delete all forms of the project, complete the following step.

- In the left pane, right-click the project and then click Delete All Forms.

Form Elements

You can add the following elements to a form.

Element	Description
Form field	Used to display and edit extracted data and to type data during manual indexing. Form fields can be of type text, amount, date, checkbox, or list.
Checkbox	Enables the user to choose between one of two possible mutually exclusive options, such as On/Off or Yes/No.
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.

Table	Used to display the table extraction result. 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.
Label	Displays a caption, usually the field name that helps the user to identify form fields, viewers or tables.
Viewer	Displays the snippet of the document from which the candidate was extracted.
Button	Uses a button to trigger actions for a script event.

Add a Form Field

To add a form field, complete the following steps.

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.

Add a Label

To add a label to a form, complete the following steps.

1. On the right vertical toolbar, click the **Add new label element** A icon.
2. On the form, click and drag to create the label.
3. In the **Change Captions** dialog box, enter a caption for the label and then click **OK**.
4. Optional. To change the caption, right-click the label and then click **Change Caption**.

About 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.

Add a Viewer

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 substeps.
   1. Select **Fixed Zone** and specify the page number in the **Page** field.
   2. 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 substeps.

   Note: This works no matter whether the field assignment is set at design time or at runtime.

   1. Select **Field**.
   2. To specify that a white border displays around the area covered by the viewer, under **Border**, specify the **Left/Right** and **Top/Bottom** values.
8. 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.

9. 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.

10. Optional. Under **Size & Position**, modify the values for size and position of the viewer on the form.

Add a Button

To add a button to a form, complete the following steps.

1. On the right vertical toolbar, click the **Add New Button** button.
2. On the form, click and drag to add the button.
3. Right-click the button and then click **Change Caption**.
4. In the **Change Caption** dialog box, enter a name and then click **OK**. For information about configuring actions triggered by buttons, see Define an action.

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 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, see “OnAction” in the *Oracle WebCenter Forms Recognition Scripting User's Guide*. To define an action, complete the following steps.

1. From the **Edit** menu, select **Define actions**.
2. To add an action, in the **Define Actions** dialog box, click **Insert**.
3. Optional. Click the **Action** field and enter a name for the action.
4. Optional. Click the **Description** field and enter a description for the action.
5. Click the **Button** field and select the button that you want to link to the action from the list.
6. Optional. To assign a shortcut to the action, click the **Accelerator** field.
7. In the **Accelerator Properties** dialog box, select **Enable this accelerator**.
8. From the **Key** list, select a key.
9. Optional. Select **Ctrl** and/or **Shift**.

   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 WebCenter Forms Recognition, the combination Ctrl + Shift is usually available to combine with any other keys.

10. Click **OK** and then click **Close**.
11. To handle the action in the `Document_OnAction` event in the script, on the toolbar, click the **Show/hide script** button.
12. Modify the script according to your needs.
13. Switch to **Verifier Test Mode** and test the action.

Delete an Element from a Form

To delete an element, complete the following steps.

1. Select the element on the form.
2. On the right vertical toolbar, click the **Delete selected elements** button.

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.

1. On the form, right-click the table and then click **Properties**.
2. In the **Properties** dialog box, complete one of the following actions.

Situation	Steps
Optimize the column width automatically according to the column content.	1. Select **Automatic Column Sizing**. 2. In the **Maximum Column Width** field, enter the value in pixels. **Note:** The user can nevertheless change the column width at runtime, but after a document change, the automatic setting is used again.
Specify a fixed width for the columns.	1. Select **Fixed Column Sizing**. 2. In the **Column Width** field, enter the value in pixels. **Note:** This value applies to all columns except the last one. The user can change the column width at runtime, but after a document change, the fixed setting is used again.
Enable the user to modify the column width at runtime.	Select **Manual Column Sizing**.

3. Optional. If you want the user to navigate through the table line by line from left to right, select **Validate Table Line by Line**.
4. Optional. If you want the user to navigate through the table column by column from top to bottom, select **Validate Table Column by Column**.
5. 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 substeps.
    - Use the arrow buttons to reorder the columns.
    - In the **Display Name** field, enter a name.
    - To protect a column from edits, select **Read-only**.

About Auto-Completion

Auto-completion helps to speed up data entry. When you start typing, the feature completes the word, suggesting the best matching items among all of the words or candidates available after OCR and format analysis.

Example

You type two or more characters of a 20-character invoice. The auto-completion feature finds the best matching candidate suggested by the Format Analysis engine and places it in that field. The auto- selected text also highlights in the original document. Select whether a single-line or a multi-line text field displays. The candidate highlights in green if the field is valid and in orange if the field is invalid. To override auto-completion, continue typing.

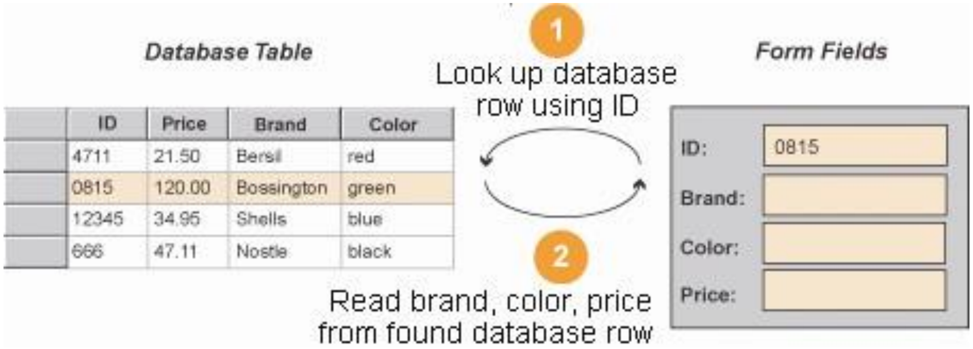
The auto-completion feature for a header field automatically selects the best candidate from the available ones. However, the viewer updates only if the candidate appears once in the document. Otherwise the viewer remains blank when auto-text completes the word for the field. This function works in the Highlight Candidates mode only.

Do not insert formatted text for auto-completion. Auto-completion does not work on formatted text and characters incorrectly read by OCR.

About 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, the following field types are available.

- Smart index field. The information related to this field must be available both in the processed documents, and in the database table. Often, a unique identifier such as "Invoice #" or "Customer #" 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 are used to fill the lookup fields.


Database Table
1
Look up database row using ID
Form Fields
2
Read brand, color, price from found database row

ID	Price	Brand	Color
4711	21.50	Bersil	red
0815	120.00	Bossington	green
12345	34.95	Shells	blue
666	47.11	Nostle	black

ID: 0815
Brand:
Color:
Price:

If used manually, the form for smart indexing should have the smart index field above the lookup fields. Otherwise, the form’s tab order disturbs the workflow.

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.

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

To configure smart indexing, complete the following steps.

Prerequisite You must have a form with at least two form fields, an existing database table that contains the corresponding columns, and an OLE DB provider for the database installed.

1. Select the form field that you want to use as a smart index field.
2. In the right vertical toolbar, click the **Define Smart Indexing** button.
3. In the **Smart indexing definition** dialog box, click the **Select Data Source** button.

   Note: The dialog displays the available fields. The smart index field highlights in yellow and is marked with a key icon.

4. In the **Data Link Properties** dialog box, on the **Provider** tab, select the appropriate OLE DB provider for the data you want to access and then click **Next**.
5. On the **Connection** tab, specify the connection properties and then click **OK**.

   Note: The properties are provider-specific. For more information, click the Help button.

6. In the **Smart indexing definition** dialog box, from the **Select database table** list, select a database table.
7. In the **Database Field** column, click a cell.

   A list containing the names of all columns of the selected table displays.

8. Map columns to the smart index field and all lookup fields.

   Note: Smart indexing fills each field that is assigned to a database column.

9. Optional. Double-click each additional field that you want to use as index field.

   Note: The search criteria are evaluated using AND joins, that means there must be a matching row for both index values.

   The corresponding fields are marked with key icons.

10. To limit the number of results, enter an entry into the **Maximum # of SmartIndex hits** field.
    Note: A database lookup using one index term may return multiple results. In interactive mode, a dialog box opens where the user can select the appropriate record.
11. Under **Further Settings**, specify where and when smart indexing should be performed. Complete any of the following substeps.
    - To perform smart indexing during the automatic extraction step by the Runtime Server, select **Use in automatic extraction**.

- To enable smart indexing during the extraction verification or the manual indexing step in Verifier, select **Use in manual field validation**.
- To perform the lookup on validation, select **Start smart indexing**. **Note** For example, the lookup is performed on validation when the user presses Enter in the index field. Otherwise, the user needs to press Alt+F12 to trigger smart indexing.
- To enable the user to check and correct the values retrieved from the database, select **Reconfirm smart indexing**. **Note** This is the recommended setting. Otherwise, the retrieved values are validated automatically.
- To accept a valid lookup result that consists of multiple values, select **Accept multiple hits**.
- To accept a valid lookup result with multiple values obtained using different index fields, select **Accept different**.

Delete the Smart Indexing Definition

To delete the smart indexing definition, complete the following steps.

1. On the form, select the smart index field.
2. In the right vertical toolbar, click the **Delete Smart Index** button.

Hide the Classes, Forms, and Fields Panes

To hide the panes, complete the following step.

- On the toolbar, click the Maximize Verification Form View button.

About Editing Text Fields

Designer 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.

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.

About the 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 assignments.

Verifier Test Mode User Interface

The Verifier Test Mode user interface provides the following panes.

Forms pane

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 pane

The lower middle pane displays a magnification of the current form field and the associated viewer. You can use it for particularly convenient viewing and editing.

Note: 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

The lower pane 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.

Verifier Test Mode Color Coding

The Verifier Test Mode uses color coding to support the quality assurance process.

- 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 a low confidence OCR character. The document area provides similar information.

- 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, you can apply document highlighting 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

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. From the **Options** menu, select **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.

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 Designer only. To modify the states, complete the following steps.

1. From the **Options** menu, select **Settings**.
2. On the **Verifier Mode** tab, under **Designer Test Mode Document Routing**, modify the states and then click **OK**.

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

Enable the Allow Batch Renaming Feature

To enable the Allow batch renaming feature, complete the following steps.

1. From the **Options** menu, select **Settings**.
2. On the **Verifier Mode** tab, select **Allow batch renaming** and then click **OK**.

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.

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.

Prepare Testing the 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 learnset.

- 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.

Test the Verification Form Layout

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.

See also [Specify 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. This icon is only visible 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.
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 substeps.
    - To inserts a new line into a multi-line field, press Ctrl + Enter.
    - To delete all character until the end of the line, press Ctrl + Del.
    - To delete the entire field content, press Ctrl + Backspace.
    - To validate the current rejected character or field and move to the next invalid character or field, press Enter. 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 and then repeat the previous steps.
10. When you checked the last field, press Enter 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.

Test the 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.

   Note: Check if the classes for which the property **Visible in correction** is cleared do not appear in the list. See also 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.

Prepare Testing the Validation Rules

To test whether your validation rules work, prepare the following document input.

- Some classified documents.

  Note: To run tests where you deliberately enter invalid input, you can use documents that have good extraction results. 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 document set with OCR problems, for instance by scanning with a lower resolution.

Test the Validation Rules

To test the validation rules, complete the following steps.

1. Load the documents.
2. Switch to **Verifier Test Mode**.

   Because the current document is classified, the extraction / indexing user interface displays.
3. To test as many violations as possible with a limited number of documents, ensure that the **Move to next document on validate** button is not pressed.
4. To activate the first field, press **Tab**.
5. Test the field validation rules as follows, and observe the messages in the status bar.

   Note: Any validation errors display in the status bar.

   - If the field contains OCR rejects, correct them and press Enter to confirm each corrected character. You should not be able to validate the field if rejects are forbidden.
   - If applicable, insert 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, insert some invalid characters and observe if they are discarded automatically.
   - If applicable, insert some characters in the wrong case and observe if they convert automatically.
   - If possible, insert some invalid input and try a forced validation by pressing **Enter** three times. If this is permitted, a message box displays. Click **No** to interrupt the forced validation.
6. Without validating the field, press **Tab** to move to the next field and then repeat the previous step.

Test Smart Indexing

To test smart indexing, complete the following steps.

Prerequisite You need access to the database used for the lookup and extracted documents 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.

1. To manually test whether the feature is working as expected, deactivate automatic smart indexing while you are testing.
2. Ensure that the Verifier Test Mode toolbar settings are in default state.
3. Load the documents.
4. Switch to **Verifier Test Mode**.

   Because the current document is classified with an invalid extraction result, the extraction / indexing user interface displays.

5. 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.
6. Run a search with the extracted value, or insert a value that returns a single result.
7. To fill the lookup fields, press Alt + F12.
8. Run a search that returns multiple results and complete the following substeps.
   1. Use a wild card search, such as using “*” to represent multiple characters, or “?” to represent a single character.

      Note: Two different results are acceptable in this case.
   2. If multiple results are forbidden by the **Maximum # of SmartIndex hits** field setting, a dialog box displays from which you can select the appropriate record.

      The lookup fields are then filled accordingly.
   3. If multiple results are allowed, the lookup fields are filled with multiple values.

      Note: You need multiline fields to display the results properly.
9. Run a search that returns no results and check if the corresponding error message displays.
10. Reactivate smart indexing.

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.

Situation	Steps
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**.

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, see the *Oracle WebCenter Forms Recognition Verifier User's Guide*. To

insert large blocks of text, complete the following steps.

1. In the document, drag a rectangle around the required text.
2. If necessary, adjust the rectangle size.
3. To copy the text from the rectangle to the field or table cell, perform one of the following steps.
    - Drag the rectangle to the required field or table cell.
    - Double-click the rectangle.

About Testing Table Analysis

The preparations for table analysis test are essentially the same as for general extraction tests. The following list 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 using scripts

The most basic test checks the presence of the table field. If some table fields are missing, you need 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.

Test the Table Analysis

To test the table analysis, complete the following steps.

Prerequisite You need some classified or extracted documents and separate batches for all document classes that use common table settings.

1. Load the documents.
2. Switch to **Verifier Test Mode**.

   Because the current document is classified, the extraction / indexing user interface displays.
3. Ensure that the **Verifier Test Mode** toolbar settings are in default state.
4. The table field provides additional highlighting options. Complete the following substeps.

   Note: Valid areas appear in green, invalid areas in red.

   - Click the upper-left corner of the table field to highlight the entire document table.
   - Click a column label in the field area to highlight the document column.
   - Click a row label in the field area to highlight the document row.
   - Click a cell in the field area to highlight the document cell.
5. Use the highlighting options to complete the following substeps.
   1. Check whether the top and bottom of the table are detected correctly.
   2. Check for default columns.

      Note: They must be available in valid tables. If a default column is missing, the corresponding table must be invalid.
   3. Check for mandatory cells.

      Note: They must be filled in valid tables. If a mandatory cell is empty, the corresponding table, column, and row must be invalid.
   4. Check whether column formats are applied correctly.

      Note: If cells have an incorrect format, the corresponding table, column, and row must be invalid.
6. Check whether forced validation of invalid tables is possible.

Correct the Table Structure

To correct the table structure, complete the following steps.

1. To correct the column structure of a table, in the left pane, complete any of the following substeps.
   - To clear all data for a column, right-click the column header and then click **Unmap Column**.
   - To exchange the position of the current column, right-click the column header, click **Swap Column**, and then select the column you want to exchange the current column with.
2. To correct the row structure of a table, in the left pane, complete any of the following substeps.
   - To insert an empty row above the current one, right-click the row number and then click **Insert Row**.
   - To delete a row, right-click the row number and then click **Delete Row**.

   - To append an empty row at the bottom of the table, right-click any row number and then click **Append Row**.
   - To merge rows, select the required rows, right-click any row number and then click **Merge Rows**.
3. Type the correct values in the cells and then press Enter to validate the table.

About Correcting Table Cells

You can correct the content of table cells by appending or replacing strings selected from the document.

A candidate is a string that matches the learned words for that cell. It highlights in green when you select it after selecting the cell. A string that does not belong to any candidates for that cell highlights in orange when selected. The append feature takes the current word behind the candidate and appends the cell text. It places the text in the best location, either right or left of the word.

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

Append or Replace Text in a Table Cell

To append or replace text in a table cell, complete the following steps.

1. In the toolbar, select the **Highlight all candidates** button.
2. To select a cell, complete one of the following substeps.
   - In the left pane, select the required table cell.
   - In the document, right-click a candidate and then click **Select Cell**.
3. Complete one of the following actions.

Situation	Steps
The word is a candidate.	- To append the selected word in the cell, in the document, right- click the required word and then click **Append Cell Text by Word**. - To replace the cell content, in the document, double-click the required word.
The word is not a candidate.	- To append the selected word in the field, right-click the box and then click **Append Cell Text by Word**. - To replace the field content with the selected word, right-click the box and then click **Replace Cell Text by Word**, or press Ctrl and click the word.

What are Regular Expressions?

Regular expressions are a generalized pattern language that WebCenter Forms Recognition uses for format analysis. With regular expressions, the engines can identify strings that match a specified pattern 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.

Literal Characters in Regular Expressions

Regular expressions use the following character types.

Literal characters

Normal text characters, such as letters or numbers, and all symbols that are not used as operators.

Special characters

The following symbols that can be used as operators.

*	-	[	]	^	{	}	\	.

To treat a special character as a literal character and not as an operator, add a preceding backslash or enclose the symbol with square brackets.

Examples

- *
- [*]

Operators in Regular Expressions

Operators consist of one or more special characters.

Operator	Description	Example
Any literal character	Match-self operator This operator is represented by any literal character. Its definition is included for completeness only. It matches the character itself.	t matches only the string t. It does not match the string tt.
This operator is not represented by any character.	Concatenation operator It concatenates two regular expressions a and b. b is simply placed after a. The result is a regular expression that will match a string if a matches its first part and b matches the rest.	xy matches xy.
.	Match-Any-Character operator This operator is represented by a period. It matches any single printing or non-printing character.	a.b matches any three-character string beginning with a and ending with b, such as axb, a5b, or a%b.

`*`	Match-Zero-or-More operator This operator is represented by an asterisk. The "*" operator is always used following a literal character or a special character. It repeats the smallest possible preceding regular expression as many times as necessary to match the pattern, including one repetition and zero repetitions.	fo* matches for example f, fo, and foo.

`{count}`
`{min,}`
`{min, max}`
	Interval operators
	The open interval operator is represented by "{".
	The close interval operator is represented by "}".
	Interval operators repeat the smallest possible preceding regular expression a specified number of times.
		{count} matches exactly <count> occurrences of the preceding regular expression.
		{min,} matches at least <min> occurrences of the preceding regular expression.
		{min,max} matches between <min> and <max> occurrences of the preceding regular expression. The interval is invalid if <min> is greater than <max>, or any of <count>, <min>, or <max> is outside the range 0 - 256.

`[ ]` `[^ ]`	Character Class operators "[" and "]" 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 "-" range operator. A character class can contain as many individual characters as needed, and can contain multiple character ranges. 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 "^" metacharacter as the first character in the list. 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. -] closes the character class unless it is the first character. - \ quotes the next character unless it is the last character before the closing bracket. - - represents the range operator unless it is the first character after the opening bracket or the last character before the closing bracket. **Note:** Empty character classes are invalid.	
–	Range operator The range operator is represented by a hyphen: -. It specifies a range of characters when used inside a character class.	b-f matches any character between b and f: b, c, d, e, or f. A-Z matches any capital letter. a-z matches any lower case letter. 0-9 matches any number.
\	Escape operator 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.	

Example: Find an Invoice Number

Task

The invoice number contains between 6 and 10 characters and possibly includes an R or a K. In addition, the digits are sometimes separated by a slash: /.

Examples

892785/222, 9R095949, or 000000484

Solution

[0-9RK/] {6,10}

Matches 892785/222, 9R095949, or 000000484.

Explanation

[0-9RK/]

The invoice number may contain any digit from 0 to 9, the character R and K, as well as a forward slash.

{6,10}

The invoice number can be between 6 and 10 characters long.

Example: Find a Date

Task

The date consists of 2-digits each for month and day and between either 2 or 4 digits for the year. The values can be separated by either a period, a slash, or a hyphen.

Example

05-12-2000 or 05.12.2000 or 05/12/2000

Solution

[01][0-9][-./][0-3][0-9][-./][0-9]{2,4}

Explanation

[01][0-9]

The first numerical expression, the month, consists of two digits. The first digit can be 0 or 1. The second digit can be any digit between 0 and 9.

[-./]

Either a dash, a period, or a slash separates the month from the day.

[0-3][0-9]

The second numerical expression, for the day of the month, consists of two digits. The first digit can be 0, 1, 2 or 3. The second digit can be any digit between 0 and 9.

[-./]

Either a dash, a period or a slash separates the month from the year.

[0-9]{2,4}

The third numerical expression, for the year, consist of 2 to 4 digits in length and can include any digit from 0 to 9.

Example: Find a Date with a Month Name

Task

Find dates formatted with month

name spelled out. Example

May 13, 2000

Solution

[AJFMSOND][a-ceg-il-pr-vy]{2,8}[ ][0-9]{1,2}[,][ ][0-9]{4}

Explanation

[AJFMSOND]

The words may begin with the listed uppercase characters. The list includes the possible month initial letters.

[a-ceg-il-pr-vy]{2,8}[ ]

The lowercase letters that are allowed in the names of the months. There can be between 2 and 8 such lowercase characters, corresponding to May and September, respectively. The name of the month is followed by a space.

[0-9]{1,2}[,][ ]

The day of the month is represented by a 1 or 2 digit number, followed by a comma and a space.

[0-9]{4}

The year is represented as a four digit number.

Example: Find an Email Address

Task

Find an email address which includes a name of up to 20 characters, the symbol @, and a domain of up to 30 characters.

Solution

[a-z]{2,20}@[a-z.]{3,30}

Explanation

[a-z]{2,20}

The first part of the email address, the name, can be between 2 and 20 characters long and may include only lower case letters.

@

The email address includes an "at" symbol.

[a-z.]{3,30}

The second part of the email address, the domain, can be between 3 and 30 characters long and may include lower case letters and periods.

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.

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 substeps.
    - 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, enter a name for the template and then click **OK**.

Note: The template is saved internally within the project.

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 substeps.
    - 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 substeps.
    1. To update the tabs for analysis or evaluation settings below with the settings from the template, click **Copy Template**.
    2. Modify the settings on the tabs below.

    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.

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**.

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 below.
5. From the **File** menu, select **Save Project**.

Note: All fields using the template immediately work with the new settings.

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 substeps.
    1. Under **Available Templates**, select a template and then click **Rename**.
    2. In the **Rename Template** dialog box, enter the new name and then click **OK**.

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**.

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 **Show Analysis 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 EXP file.

5. In the **Save As** dialog box, in the **File name** field, enter a name and then click **Save**.

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 **Show Analysis Templates**.
3. In the right pane, on the **Analysis Templates** tab, click **Import**.
4. In the **Open** dialog box, select an EXP file and then click **Open**.
5. In the **Import Templates** dialog box, clear the check boxes for the templates you do not want to import and then click **OK**.

   Note: The program automatically recognizes the template type contained in the export file. If the type is wrong, the template is not imported.

About Script Encryption

The script encryption feature enables you to encrypt the scripts of a specific class in a project. The following limitations apply to encrypted scripts.

- You cannot encrypt scripts on project level.
- You cannot encrypt scripts containing syntax errors.
- To prevent unauthorized users from accessing the content of encrypted scripts, using the **Step Into (F8)** function is only possible two times when debugging encrypted scripts.

This limitation does not apply if the user entered the private script access key prior to debugging.

Activate Script Encryption

To activate the script encryption, complete the following steps.

1. Complete one of the following substeps.

   Note: Do not create any encryption keys on customer site.

   - For a 32-bit machine, create a registry file with the following content.

     ```
     Windows Registry Editor Version 5.00
     [HKEY_LOCAL_MACHINE\SOFTWARE\Oracle\Project
     Encryption] "Encryption Class Name"="ClassName"
     "Encryption
     Activated"=dword:00000001
     "Encryption Private
     Key"="PrivateKey"
     "Encryption Public Saving Enabled"=dword:00000000
     ```

   - For a 64-bit machine, create a registry file with the following content.

```
Windows Registry Editor Version 5.00
[HKEY_LOCAL_MACHINE\SOFTWARE\Wow6432Node\Oracle\Pro
ject Encryption]
"Encryption Class
Name"="ClassName" "Encryption
Activated"=dword:00000001
"Encryption Private
Key"="PrivateKey"
"Encryption Public Saving Enabled"=dword:00000000
```

2. Set the value for the `Encryption Class Name` key to the name of the class that contains the scripts you want to encrypt.
3. Set the value for the `Encryption Private Key` key to an arbitrary string.

   Note: The maximum character length for the key is 260.
4. Run the registry file.
5. Optional. Repeat the previous steps for any other required class.
6. In **Designer**, open the project.
7. For security reasons, we recommend saving the file with a different name. Complete the following substeps.
   1. From the **File** menu, select **Save As** to save the project with a different name.
   2. From the **File** menu, select **Save Compress**.
8. In **Windows Registry Editor**, delete the `Project Encryption` key.

   Note: Once a class is encrypted, these keys are no longer required and should be removed for security reasons.

Debug an Encrypted Script

To debug an encrypted script, complete the following steps.

1. In **Designer**, open the project.
2. In the left pane, on the **Classes** tab, right-click the project and then click **Enable Script Access**.
3. In the **Enable Script Access** dialog box, enter the `Encryption Private Key` you created in Activate script encryption and then click **OK**.
4. Wait for 5 seconds and then debug or open the encrypted script.

   Note: The script remains encrypted even though you modify and save the script.

Decrypt a Project

To decrypt all script in a project, complete the following steps.

1. Complete one of the following substeps.

   Example

   - For a 32-bit machine, create a registry file with the following content.

```
Windows Registry Editor Version 5.00
[HKEY_LOCAL_MACHINE\SOFTWARE\Oracle\Project
Encryption] "Encryption Public Saving
Enabled"=dword:00000001
```

  - For a 64-bit machine, create a registry file with the following content.

    ```
    Windows Registry Editor Version 5.00
    [HKEY_LOCAL_MACHINE\SOFTWARE\Wow6432Node\Oracle\Pro
    ject Encryption]
    "Encryption Public Saving Enabled"=dword:00000001
    ```

2. Run the registry file.
3. In **Designer**, open the project.
4. In the left pane, on the **Classes** tab, right-click the project and then click **Enable Script Access**.
5. In the **Enable Script Access** dialog box, enter the `Encryption Private Key` you created in [Activate script encryption]() and then click **OK**.
6. From the **File** menu, select **Save Compress**.

Use Breakpoints in Encrypted Scripts

Breakpoints in encrypted scripts disappear when the user stops the script. To avoid this, complete the following steps.

1. Apply a dummy classification or extraction to a document classified to the class. This ensures that all scripts are loaded. The document to debug is classified later.
2. Set the required breakpoints.

   Note: If you stop the script, you need to set the breakpoints again.

Bypass Debugging Issues

Complete the following steps when an encrypted script stops unexpectedly while debugging.

1. Switch to **Definition Mode**.
2. In the left pane, on the **Classes** tab, right-click the project and then click **Enable Script Access**.
3. In the **Enable Script Access** dialog box, enter the `Encryption Private Key` and then click **OK**.
4. On the toolbar, click the **Show/hide script** button.
5. Stop the script if it is running.
6. In **Definition Mode**, in the left pane, select the **Fields** tab.
7. Click the **Classify / analyze current document** button.
8. In **Definition Mode**, in the left pane, select the **Classes** tab.
9. On the **Classes** tab, select an encrypted class and then click the **Show/hide script** button.
10. In the script, set the required breakpoints.
11. Run classification or extraction and debug the script.

Define a Complex Sequence for Class script Dependencies

In a complex project, the project script might not be the primary script depending on other class level scripts, including one or more encrypted scripts.

Example

A project contains the following dependent classes.

1. **GlobalVariables** class script (unencrypted)
2. **UserExits** class script (unencrypted)
3. **BWPackaged** class script (encrypted)
4. **Project** script (unencrypted)
5. **Invoices** class script (unencrypted)
6. **Generic** class script (encrypted)

To define dependencies between scripts, add an `#include "ClassName"` statement in the script dependent on the **ClassName** class.

The **GlobalVariables** class script is the primary script and therefore does not require an `#include` statement.

The **Invoices** and **Generic** class scripts do not require modifications.

Note: You can only add one `#include` statement per script.

1. Open the **UserExits** class script and add the following line at the bottom of the script.

   Example

   ```
   #include "GlobalVariables"
   ```

2. Open the **BWPackaged** class script and add the following line at the bottom of the script.

   Example

   ```
   #include "UserExits"
   ```

3. Open the **Project** script and add the following line at the bottom of the script.

   Example

   ```
   #include "BWPackaged"
   ```

Regional and Currency Settings

The following table lists settings for regions 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, €