Course Guide

IBM Datacap 9.0.1: Datacap Studio

Course code F255 ERC 2.2

IBM Training

December 2019 edition

Notices

This information was developed for products and services offered in the US.

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not grant you any license to these patents. You can send license inquiries, in writing, to:

IBM Director of Licensing
IBM Corporation
North Castle Drive, MD-NC119
Armonk, NY 10504-1785
United States of America

INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some jurisdictions do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM websites are provided for convenience only and do not in any manner serve as an endorsement of those websites. The materials at those websites are not part of the materials for this IBM product and use of those websites is at your own risk.

IBM may use or distribute any of the information you provide in any way it believes appropriate without incurring any obligation to you.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to actual people or business enterprises is entirely coincidental.

Trademarks

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the web at “Copyright and trademark information” at www.ibm.com/legal/copytrade.shtml.

Contents

Trademarks

The reader should recognize that the following terms, which appear in the content of this training document, are official trademarks of IBM or other companies:

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corp., registered in many jurisdictions worldwide.

The following are trademarks of International Business Machines Corporation, registered in many jurisdictions worldwide:

DB2®	FileNet®	Initiate®
Tivoli®	WebSphere®	

Microsoft and Windows are trademarks of Microsoft Corporation in the United States, other countries, or both.

Other product and service names might be trademarks of IBM or other companies.

Course description

IBM Datacap 9.0.1: Datacap Studio

Duration: 2 days

Overview

This course shows you how to create and configure a Datacap application using Learning Template and then to enhance it using Datacap Studio. You identify when to use a Learning Template rather than a Form Template for your application.

You work with a fully functioning IBM FileNet Content Manager system configured with IBM Content Navigator and IBM Datacap 9.0.1 to practice the skills that are required to implement and configure data capture solutions.

Audience

- Application Builders who are responsible for building Datacap applications
- Anyone who needs to know how to rapidly develop a capture solution and enhance the application using the full capability of the IBM Datacap suite.

Prerequisites

The following courses or equivalent knowledge are required:

- IBM Datacap 9.0.1: Introduction (F251)
- IBM Datacap 9.0.1: FastDoc Server Mode (F253)
- IBM Datacap 9.0.1: FastDoc Multi-Page Document (F254)

Skills taught

Upon completion of this course, participants will be able to:

- Work with Datacap Studio
  - Explore Rulemanager, Zones, and Test views in Datacap Studio
- Create an application and configure in FastDoc
  - Create an application based on the Learning Template
  - Configure Batch Structure (Document Hierarchy)
  - Setup Image Enhancement
  - Configure Page recognition

- Enhance the application in Datacap Studio
  - Create the Locate rules to locate fields with keyword search
  - Locate fields with intellocate (automatic in the learning template)
  - Configure the field Validation rules to validate the fields on the pages
  - Extend the Locate and Validation rules to handle more page types
- Export to IBM FileNet Content Manager repository
  - Convert document images to pdf format
  - Configure Export Rules for IBM FileNet Content Manager
- Application debugging and Test the task profiles
  - Use the Test tab in Datacap Studio for application debugging
  - Locate and use Datacap Logs to troubleshoot Datacap error conditions

Course outline

- Introduction to Datacap Studio
- Create an application based on Learning Template
- Enhance the application in Datacap Studio
- Create Locate and Zone rules
- Create Validate rules
- Expand the Locate and Validation rules
- Document export to corporate data repository
- Application Debugging and Logs

Agenda

Note

The following lesson durations are estimates, and might not reflect every class experience.

Day 1

Unit 1.Datacap Studio Introduction

- (0:30) Lesson 1 - Datacap Studio - Rulemanager
- (0:30) Lesson 2 - Datacap Studio - Zones and Test views

Unit 2.Basic Learning Application

- (01:00) Lesson 1 - Create a Learning Template application
- (01:00) Lesson 2 - Configure a Learning Template application
- (01:30) Lesson 3 - Create locate rules
- (01:30) Lesson 4 - Create validate rules

Day 2

Unit 2.Basic Learning Application

- (01:30) Lesson 5 - Expand the locate rules
- (01:00) Lesson 6 - Expand the validate rules
- (01:30) Lesson 7 - Export to an IBM FileNet repository

Unit 3.Debug and Test

- (01:00) Lesson 1 - Application debugging
- (01:00) Lesson 2 - Datacap Studio Test tab

Unit 1. Datacap Studio Introduction

Estimated time

01:00

Overview

This unit introduces you to the capabilities Datacap Studio Application Building capabilities.

How you will check your progress

- Successfully complete the activities in the Student Exercises book.

References

IBM Knowledge Center

http://www.ibm.com/support/knowledgecenter/SSZRWV_9.0.1/com.ibm.datacaptoc.doc/datacap_9.0.1.htm

IBM Training

Unit Objectives

- After completing this unit, you should be able to:
- Describe the components in Datacap Studio for application development process
- Use the Datacap Studio

Figure 1-1. Unit Objectives

Lesson 1.1. Datacap Studio – Rulemanager

Figure 1-2. Datacap Studio – Rulemanager

IBM Training

Lessons

- Datacap Studio – Rulemanager
- Datacap Studio – Zones

Figure 1-3. Lessons

IBM Training

Why is this lesson important to you?

- As an Application Developer, you build and deploy applications with the Datacap capture system.
- To do these tasks effectively, you must be familiar with the Datacap Studio interface.

Figure 1-4. Why is this lesson important to you?

IBM Training

Panel Organization within Datacap Studio

- Three Main Tabs
  - Rulemanager, Zones, and Test Tabs.
- Rulemanager Tab
  - Contains 5 panels where you view document structures, rulesets, rules, functions, and task profiles.
  - Can also be used to extend application created in Datacap FastDoc.
- Zones Tab
  - Contains 4 panels where you add and define fingerprints and view and modify properties of selected objects.
- Test Tab
  - Contains 8 panels in which you can view information and properties of batches, jobs, documents and rulesets.
  - Is a test environment that allows you to debug your application, set breakpoints, step through and skip pieces of your application, and lots more functionality.

Figure 1-5. Panel Organization within Datacap Studio

Panel Organization within Datacap Studio

Datacap Studio has three views that you select with the three tabs across the top of the window.

- Rulemanager
- Zones
- Test

The Rulemanager View

This pane is the primary application development area. The default layout of the Rulemanager view for Datacap Studio window has four panes. One of these panes has two functions selectable by tabs at the top of the pane. You can customize the pane layout by removing or adding panes or by moving them to different locations. The Rulemanager panes are:

- Document hierarchy
- Rulesets
- Action library
- Task profiles
- Properties

The Zones View

Contains 4 panels where you add and define fingerprints and view and modify properties of selected objects.

The Test Tab

Contains 8 panels where you can view information and properties of batches, jobs, documents and rulesets.

It also functions as a test environment that allows you to debug your application by setting breakpoints,

stepping through and skipping blocks of your application, along with many more debugging capabilities.

Document hierarchy

The document hierarchy describes the structure of the documents that your application is designed to handle.

The levels within the hierarchy are:

- Batch
- Document
- Page
- Field.

Rulesets

- A ruleset consists of one or more rules.
- Rulesets are assigned to tasks in the task profiles.
- Rules are assigned to process specific objects in the document hierarchy (for example, to analyze each page and identify its type).

There are two Types of Ruleset:

- Conventional rulesets
- Compiled or UI rulesets

Task profiles

A workflow consists of a series of tasks and defines a way to process documents. Datacap applications can include multiple workflows. Each task is linked to a task profile and the task profile determines the order that tasks are processed.

Actions library

Provides access to the complete library of prebuilt actions. An action Is the smallest component of a workflow. It is the equivalent of an instruction in programming terms. To get help on an action, select the action and click Information.

Properties

Shows the properties for the selected document hierarchy or ruleset object. If the corresponding pane is locked for editing, you can also modify existing properties, including specifying action parameters.

IBM Training

Datacap Studio Overview

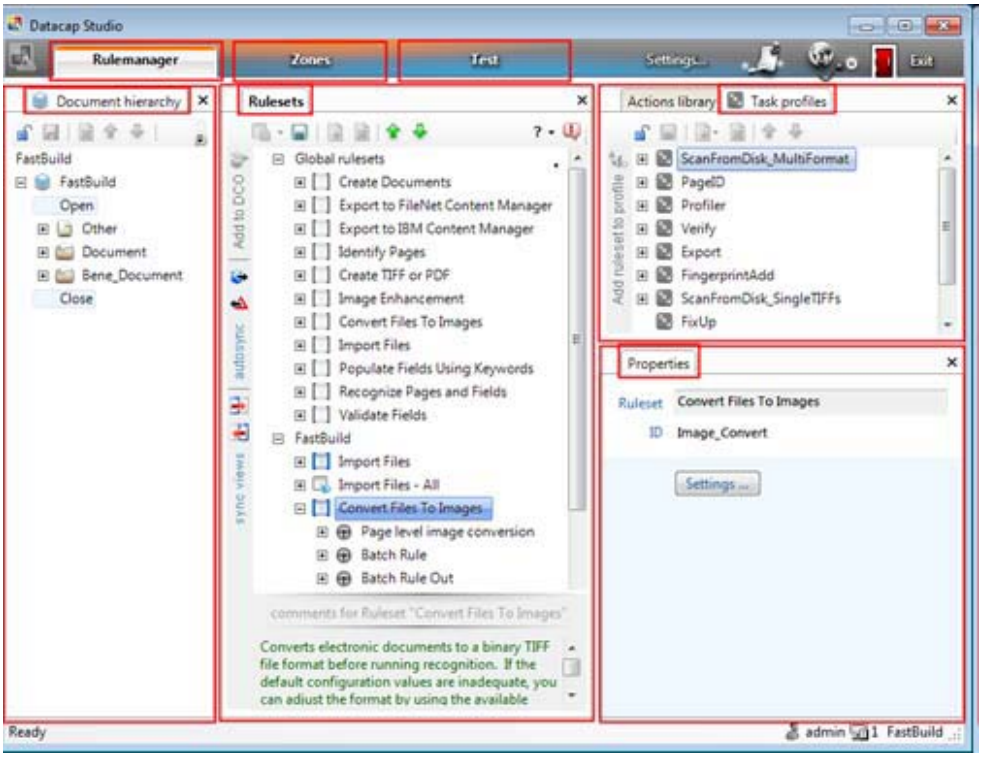


Figure 1-6. Datacap Studio Overview

Datacap Studio overview

- Datacap Studio is a full function rich application development environment for Datacap Capture.
- It also includes an Application Wizard that gives you a head start on application development by generating a basic application framework, complete with the supporting folders.
- It provides the tools that you need to develop and test your application.
- Any application that is configured in the FastDoc rapid development environment can be developed further in Datacap Studio to achieve functions that are not supported in FastDoc.
- Up to Datacap 8.1 it was the only application development tool.
- In Datacap 8.1, FastDoc was a limited functioning rapid application development tool.
- In Datacap 9.0, FastDoc abilities are vastly expanded.

Start Datacap Studio

To open a sample application in Datacap Studio:

1. Click Start > All Programs > IBM Datacap Developer Tools > Datacap Studio.
2. In the Select Application window, select one of the existing sample applications (for example, FastForm) and click Next.
3. In the Taskmaster Login window, make sure that the NT authentication check box is not selected.
4. Enter User ID:, Password:, Station ID and click Finish. (Default Credentials User ID = admin, Password = admin, : Station ID = 1,

IBM Training

Rulemanager View

- Document hierarchy
  - Describes the structure of the documents your application is designed to handle.
- Rulesets
  - Rulesets are assigned to tasks in the Task profiles.
  - Rules are assigned to process specific objects in the document hierarchy.
- Task profiles
  - Lists the workflow of tasks. Determine the order that rules are processed while processing a task.
- Actions library
  - Provides access to the complete library of prebuilt actions.
- Properties
  - Shows the properties for the selected document hierarchy or ruleset object.

Figure 1-7. Rulemanager View

Document hierarchy

The document hierarchy describes the structure of the documents that your application is designed to handle.

The levels within the hierarchy are **Batch**, **Document**, **Page,** and **Field**.

Ruleset

A ruleset consists of one or more rules and is assigned to tasks in the task profiles.

Rules are assigned to process specific objects in the document hierarchy (for example, to analyze each page and identify its type).

There are two types of rulesets:

- Conventional rulesets
- Compiled or UI rulesets

Task profiles

A workflow consists of a series of tasks and defines a way to process documents. Datacap applications can include multiple workflows. Each task is linked to a task profile and the task profile determines the order that tasks are processed.

Actions library

A complete library of prebuilt actions. An action Is the smallest component of a workflow. It is the equivalent of an instruction in programming terms. To get help on an action, select the action and click Information.

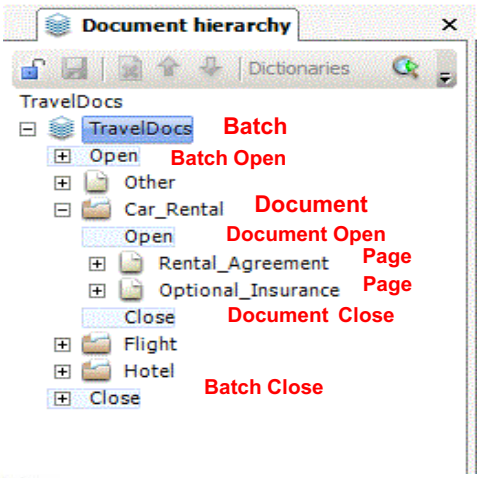
Properties

Shows the properties for the selected document hierarchy or ruleset object. If the corresponding pane is locked for editing, you can also modify existing properties, including specifying action parameters.

IBM Training

Document Hierarchy (DCO)

- Component levels are:
  - Batch
  - Document
  - Page
  - Field
- There are open and close nodes for each component.



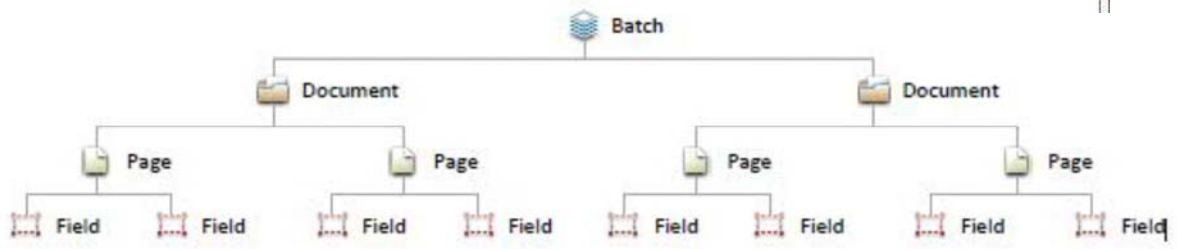


Figure 1-8. Document Hierarchy (DCO)

The images on the slide illustrate:

- Document Hierarchy view.
- A graphical representation of the Batch Structure Batch, Document, Page, Field.

Document Hierarchy

The Document hierarchy is also known as the DCO. In FastDoc the DCO is also known as the Batch Structure.

The component levels within the hierarchy are:

- **Batch**

  At the top of the document hierarchy, is the batch, which refers to all pages of all document types.

- **Document**

  Beneath the batch level, the document hierarchy defines:

  - The document types that your application can process. You might have only one type, or you might have multiple types.

Example: TravelDocs application processes car rental documents, airline ticket, humanitarian relief fund expense forms, and donation receipts.

- **Page**
  - Other page type

    There is always a page type of Other associated with the batch. This page type is used to temporarily assign a page type to all incoming pages until each page is identified as being one of the recognized pages in the application.
  - Custom pages that are associated with a defined document.

    The page types define the pages within each document type. Each document might have only one page type, or it might have multiple types. The page definitions define the number and order of pages within each document type. Pages can be required or optional.

    Example: The car rental document includes the rental agreement page and the optional insurance page, while the flight document has only an air ticket page.

    Example: A car rental document has a maximum of two pages. The rental agreement page is required and must come first. The insurance coverage page is optional.
- **Field**
  - The field nodes of the DCO are the data fields within each page type. Data fields can be required or optional.
  - Data fields configured for extracting data from the images are mapped to the fields defined in the DCO.
- **Open and Close nodes.**

  Each component has an Open and a Close node in the hierarchy. The Open and Close nodes are used to assign processing rules to the beginning or the end of each component.

IBM Training

Rules and Rulesets

- **Rule**
  - Entity that is tied to a DCO object.
  - An ordered set of functions that processes an object until a function is evaluated as True or all the other functions are exhausted.
  - Reusable and extensible.
- **Ruleset**
  - A set of rules that is used to do a particular type of processing on various DCO objects, i.e. the Validate Ruleset.
- **Task Profile**
  - A task profile is made up of a number of rulesets that are arranged in a particular sequence to produce the desired task processing results.

Figure 1-9. Rules and Rulesets

Rule

A rule is the definition of some processing to be done and is tied to a DCO object. It is made up of an ordered set of functions that processes an object until a function is evaluated as True or all the other functions are exhausted. It is reusable and extensible.

Rulesets

A ruleset consists of one or more rules and is assigned to tasks in the task profiles. Rules are assigned to process specific objects in the document hierarchy (for example, to analyze each page and identify its type). There are two types of rulesets in Datacap 9.0:

- Conventional rulesets
- Compiled or UI rulesets

Task profiles

A workflow consists of a series of tasks and defines a way to process documents. Datacap applications can include multiple workflows. Each task is linked to a task profile and the task profile determines the order that tasks are processed.

IBM Training

Rules and Rulesets

- Rulesets and Rules:

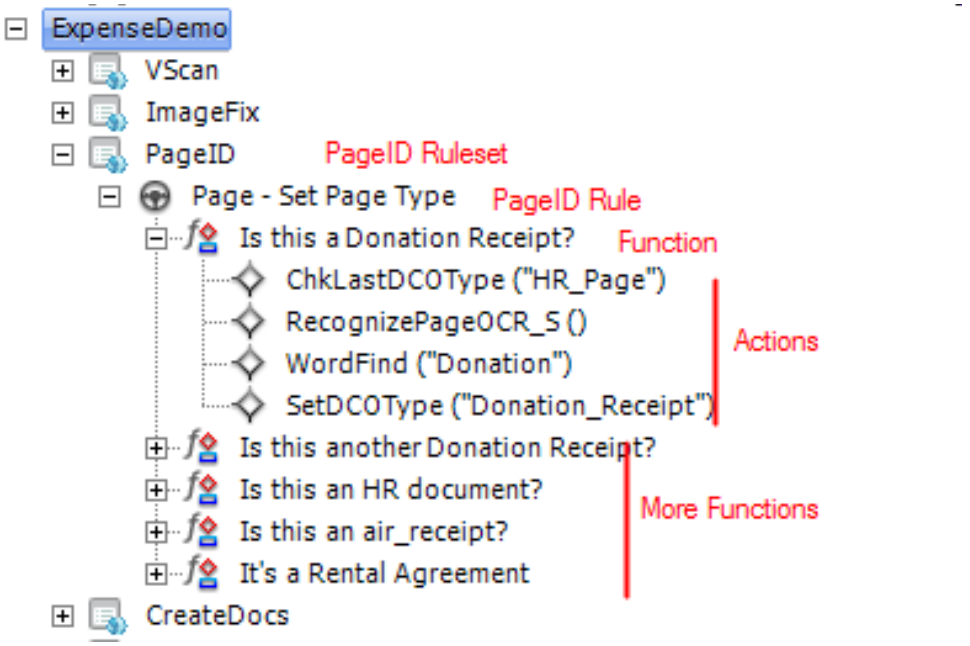


Figure 1-10. Rules and Rulesets

The image on the slide illustrates a sample ruleset, showing the structure of Batch, Rulesets, Rules, Functions, Actions.

Rulesets

A ruleset consists of one or more rules. In the following example, the default PageID ruleset has one rule: Page – Set page Type. You can see the rules that are associated with each ruleset in the Rulesets pane on the Datacap Studio.

Rulesets view.

Rules

Rules are assigned to process specific objects in the document hierarchy (for example, to analyze each page and identify its type). However, you define the rule by the programmed functions and actions that you put in it. The default PageID rule consists of five functions and multiple actions in each function, as shown in the screen capture.

The PageID function first runs the ChlLastDCOType(“HR_Page”) action. If ChkLastDCOType is successful (returns True), the function runs RecognizePage(OCR_S().

If ChkLastDCOType fails (returns False), the function fails and Datacap Capture runs the next function within the rule.

If processing reaches the end of a rule and no error is encountered then the true condition is returned and Datacap continues to process the next rule for the current object.

Add Rules to the DCO.

See Rule mapping in subsequent slides.

IBM Training

Actions and Functions

- Actions
  - Fundamental building block from which rules are built.
  - Built with .Net functions or VB script functions.
  - Always return True or False to the Rulerunner service.
  - A group of actions forms a function.
- Functions
  - Ordered list of actions that run until one of the actions returns the value of False.
  - A rule can have multiple functions that will be executed in the listed order until one of them completes, or all functions have failed.
  - If all functions fail, the rule did not execute successfully.

Figure 1-11. Actions and Functions

Actions

Actions are the fundamental building blocks that are used to define rules. They always return True or False to the Rulerunner service and may perform other types of actions. A group of actions forms a function.

Functions

A function is an ordered list of actions that will run until one of the actions returns a value of False. Rules are made up of functions that will execute in the listed order until one of the functions completes or all functions have failed.

Actions library

Provides access to the complete library of prebuilt actions for a particular type of functionality. An action is the smallest component of a workflow. It is the equivalent of an instruction in programming terms. To get help on an action, select the action and click Information.

Properties

Shows the properties for the selected document hierarchy or ruleset object. If the corresponding pane is locked for editing, you can also modify existing properties, including specifying action parameters.

IBM Training

Global and Application Rulesets

- Global rulesets:
  - RRS folder
- Application rulesets:
  - Application folder

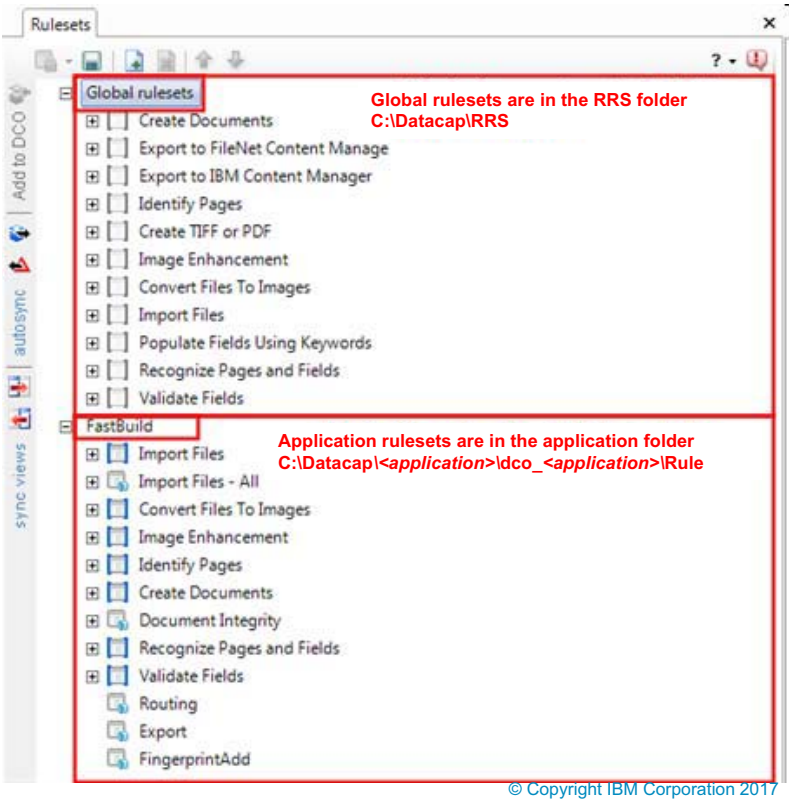


Figure 1-12. Global and Application Rulesets

The image on the slide illustrates the view of the Datacap Studio Ruleset pane :

- It shows the Global rulesets in the upper section
- It shows the Application rulesets in the lower section

Global rulesets:

- Are a pool of built-in rulesets included with the Datacap release.
- Are in the C:\Datacap\RRS folder.
- Are typically but need not necessarily be compiled (UI) rulesets.
- They are ruleset templates and cannot be configured while in the Global ruleset group. Icon is gray (Not configurable)
- A copy must be placed into the application before the application copy can be configured.
- Developers can copy customized application rulesets into the RRS folder. This makes them global and available for other developers to share.

Application rulesets:

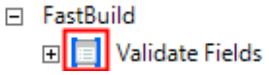
- Are the rulesets that are selected to build the application.
- Are in the C:\Datacap\<*application*>*dco_*<*application*>*rules* folder.
- Might be a copy of a global ruleset.
- Can be conventional or compiled rulesets.
- Icons are not gray but have color. (Configurable)
- For compiled (UI) rulesets, you can access and configure the properties that are configurable from through the FastDoc UI interface.
- Compiled rulesets cannot be locked and structurally modified through Datacap Studio

Note: The Global rulesets that are in the RRS folder cannot be configured because they are templates. They must be copied into the application folder by doing a copy and paste before they become configurable.

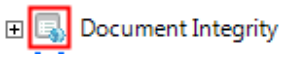
IBM Training

Ruleset Types

- Compiled UI rulesets:
  - Have a user interface to make parameter configuration easier.



  - Have a symbol that is bracketed by vertical bars.
- Conventional rulesets:
  - Have no user interface to make parameter configuration easier.



  - Have the traditional ruleset icon.

Figure 1-13. Ruleset Types

The images on the slide illustrate:

- The icon used to represent compiled UI rulesets
- The icon used to represent conventional rulesets

Compiled rulesets:

- Are self-contained application building blocks.
- Reduce the expertise that is needed to create applications.
- Are used in the built-in Forms and Learning templates.
- Make applications more consistent and easier to maintain.
  - Done by standardizing how core functions are implemented.
- Can be easily included into an application by an Application Builder.
- Have two files that represent the ruleset.
  - ValidateFields.Rul.dll - The .Rul.dll files contain the rule, function, and action definitions.
  - ValidateFields.Rul.dll.config - The .Rul.dll.config files are XML files that contain the settings for the user interface.

- Configuration can be done through Settings either with FastDoc or Datacap Studio.

Conventional rulesets:

- Rulesets that are created in legacy applications that are created before Datacap 9.0.
- Custom rulesets that are created with Datacap Studio to do specific functions that cannot be accomplished with the built-in Compiled (UI) rulesets.
- Have access to hundreds of actions that can be used to create powerful capture solutions.
- Can also be used in the built-in Forms and Learning templates.
- Can be easily included into an application with FastDoc when it is created and configured in Datacap Studio.
- Have one file that represents the ruleset.
  - ValidateFields.Rul.dll - The .Rul.dll files contain the rule, function, and action definitions.
- Configuration can be done only by setting the action parameters with Datacap Studio.

IBM Training

Rule Mapping

- Rule mapping can be done only by using Datacap Studio.
- There are two methods of doing rule mapping:
  - Legacy mapping method:
    - Rule mapping data in legacy applications is stored in the dco_<application>.xml file.
  - Datacap 9.0 mapping method:
    - Rule mapping data in Datacap 9.0 applications is stored in the <ruleset>.Rul.dll.config file.
    - This method makes it easier to share rulesets with other applications or between developers.
    - This method greatly simplifies application management.
    - Manual configuration or need to develop rules to do application management now happens automatically.
    - Mappings reduce the need for learning how to configure routing and other more technical issues as Datacap 9.0 is enhanced.

Figure 1-14. Rule Mapping

Legacy rule mapping method:

To map a rule to a DCO object:

1. Select the object in the Document hierarchy pane that you must add the rule to.
2. Then, select the rule to add.
3. Next, click the Add to DCO link on the left of the Ruleset pane.
4. The default behavior is to add the rule to the Open node of the selected DCO object.
5. To add a rule to the Close node, you must explicitly select the Close node in the Document hierarchy pane before clicking Add to DCO.

Datacap 9.0 mapping method:

To map a rule to a DCO object:

1. Expand a compiled (UI) ruleset
2. Select the rule that you want to map.
3. The mapping rule options are shown in the Properties pane.
4. Expand either “Run rule at start of …” or “and at end of …” view.

5. Select the Object type that you want to map the rule to.
6. If you want to map to a specific object of the selected type, click the + symbol and select the specific object.

IBM Training

Properties tab

- View object properties
- In general the properties pane shows the following:
  - Object type
  - Object name
  - Object properties, static and configurable. (If applicable)

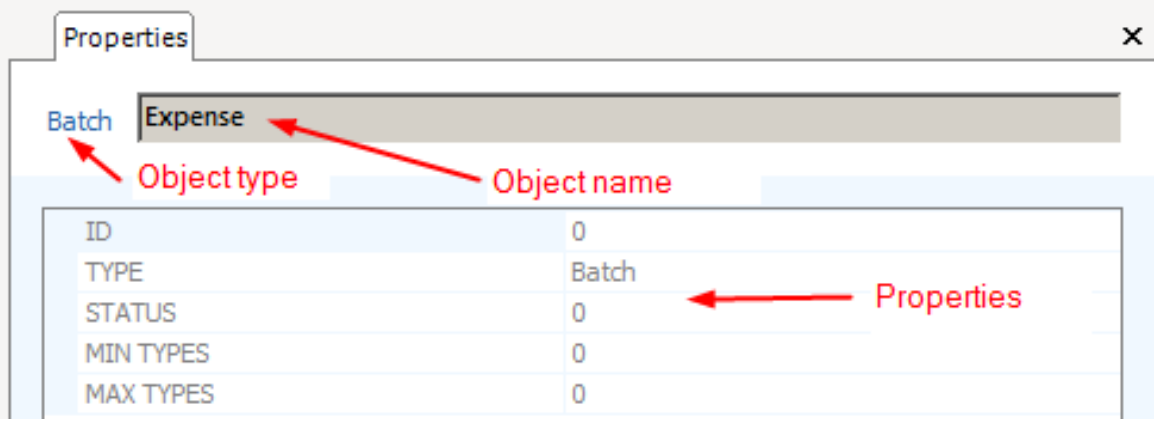


Figure 1-15. Properties tab

The image on the slide illustrates the view of the Properties tab on the Datacap Studio Rulemanager window.

Properties

- Shows the properties for the selected document hierarchy or ruleset object. If the corresponding pane is locked for editing, you can also modify existing properties, including specifying action parameters.
- Each object type in the Document hierarchy and the Rulesets pane have a unique representation in the Properties pane.
- Document hierarchy object types are:
  - Application
  - Batch
  - Document
  - Page
  - Field

- Ruleset object type properties are:
  - Application
  - Ruleset
  - Rule
  - Function
  - Action
- On the following slides you look closer at the properties view for Ruleset, Rule, Function and Action objects.

Ruleset object properties tab

- Configure UI Settings
  - For compiled (UI) rulesets, the Settings … control opens the UI interface for setting the Ruleset configuration options.
  - The Settings UI view is the same view seen in FastDoc for setting Ruleset configuration options.

Figure 1-16. Ruleset object properties tab

The image on the slide illustrates Properties view of the Datacap Studio window when a ruleset in Selected in the ruleset pane. It shows the settings control that opens the Compiled UI interface configuration options.

IBM Training

Rule object properties tab

- Map Rule to DCO objects

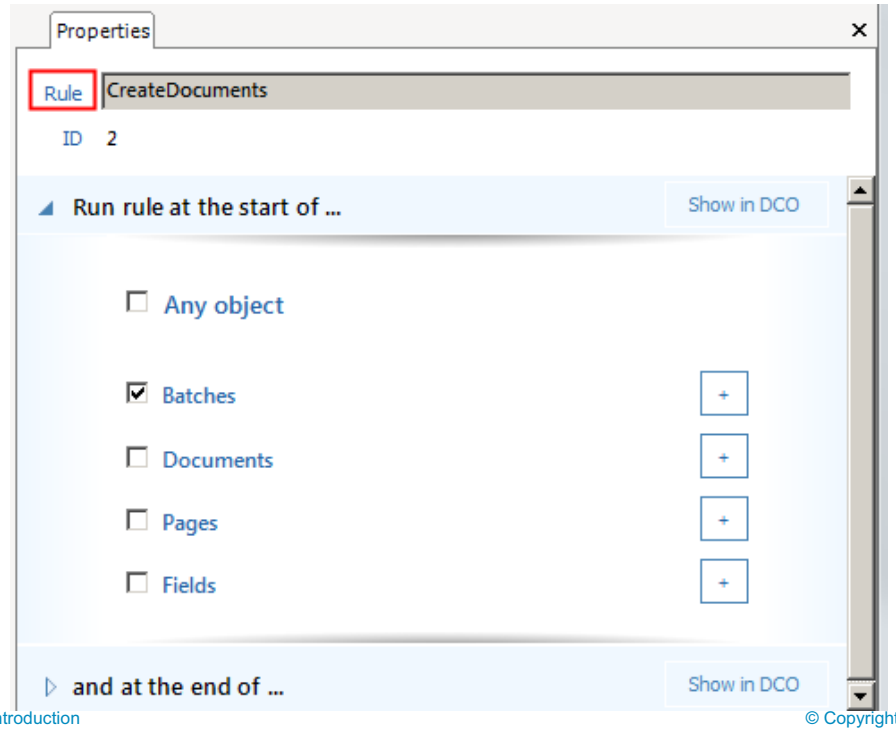


Figure 1-17. Rule object properties tab

The image on the slide illustrates Properties pane of the Datacap Studio window when a rule is selected in the Ruleset pane. It shows the controls used to associate rules with the DCO or Batch Structure objects.

Conventional Rule Mapping Method

- Conventional Ruleset rules are mapped using the pre 9.0 method.
  - Select the Rule in the Ruleset pane, select the DCO object in the Document hierarchy pane, and click the Add to DCO control.
  - Using the conventional method the mappings are visible in the Open or Close node for DCO object.

Datacap 9.0 rule Mapping Method

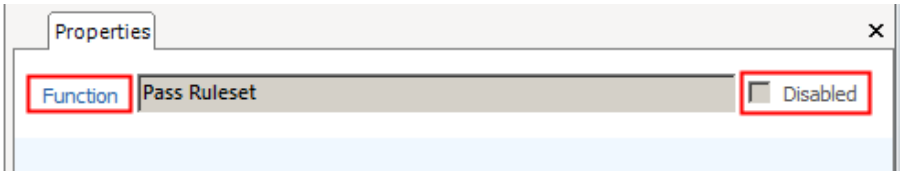
- Only for Compiled (UI) ruleset rules.
  - Rule mapping can only be done this way (as shown in the slide graphic) for Compiled (UI) Rulesets.
- Mapping at beginning or end of object rule processing.
  - Mapping configuration done under ‘Run rule at the start of …’ is equivalent to mapping a rule to the Open node in the conventional method.

  - Mapping configuration done under ‘and at the end of …’ is equivalent to mapping a rule to the Close node in the conventional method.
- Mapping to all objects or all objects a a specific type.
  - The ‘Any object’ check box causes the selected rule to associate with all DCO objects.
  - The ‘Batch’ check box causes the selected rule to associate with all batch level objects.
  - The ‘Documents’ check box causes the selected rule to associate with all document level objects.
  - The ‘Page’ check box causes the selected rule to associate with all page level objects.
  - The ‘Field’ check box causes the selected rule to associate with all Field level objects.
- Mapping to a specific object of a specific type.
  - To map to a specific object instead of all objects of one type, click the + symbol and select the DCO object from the object list that you are presented with.

IBM Training

Function and Action object properties tab

- Disable the Function



- Disable the Action and configure action parameters

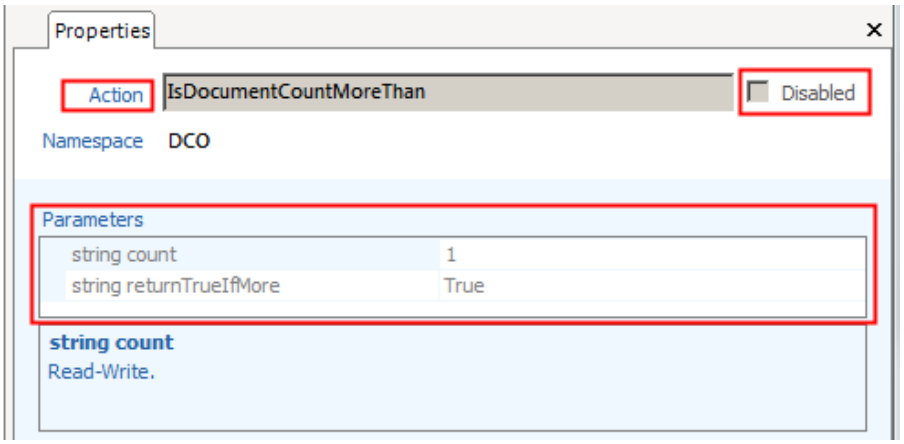


Figure 1-18. Function and Action object properties tab

The images on the slide illustrate the properties pane for functions and for actions that allow for disabling the selected object.

Disable Functions and Actions.

- The disable Function and Disable action options are to enable the activation or deactivation of UI Ruleset options.
- It can also be used for debugging purposes.

Set Action Parameters.

- The parameters that are required by the Actions are provided in the parameters area of the Properties pane.

IBM Training

Action object properties tab

- View object properties

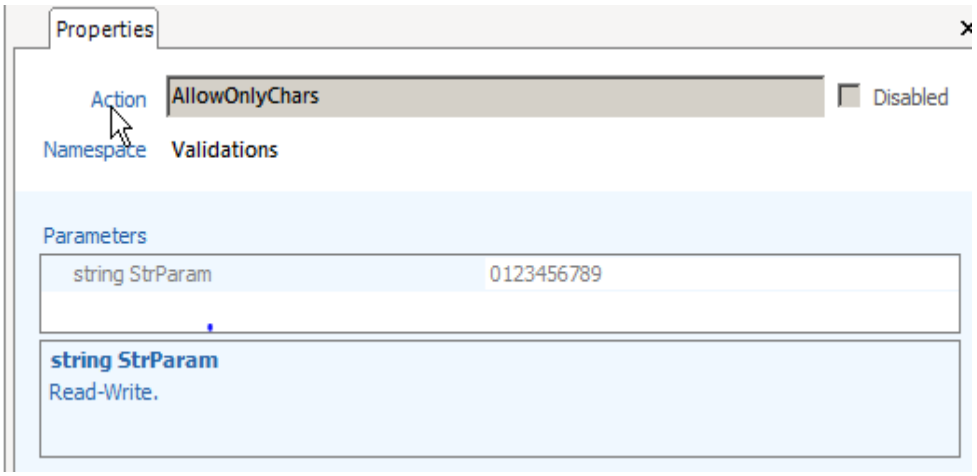


Figure 1-19. Action object properties tab

Properties

Shows the properties for the selected document hierarchy or ruleset object. If the corresponding pane is locked for editing, you can also modify existing properties, including specifying action parameters. The properties pane is where you go to set the parameters that are passed to actions.

IBM Training

Task Profiles

- Tasks and Rulesets

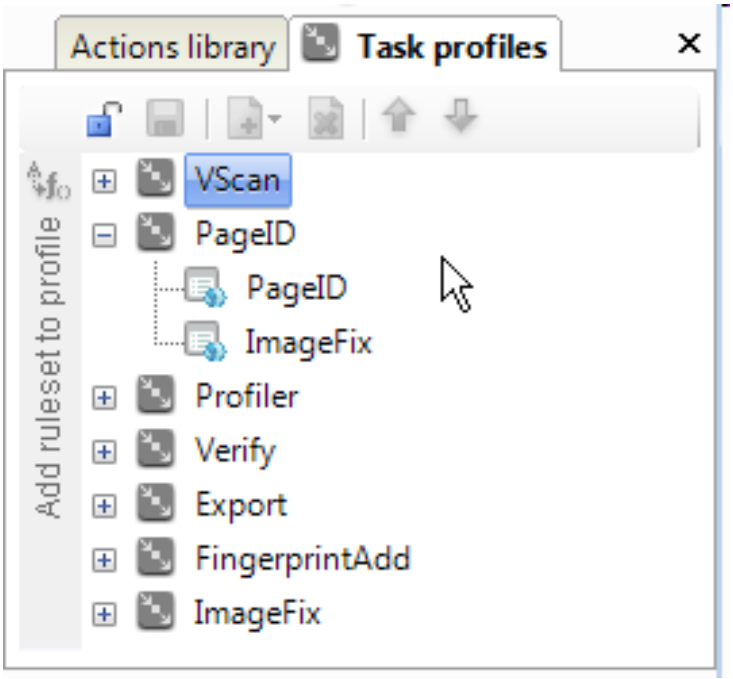


Figure 1-20. Task Profiles

The image on the slide illustrates the task Profiles pane on the Datacap Studio Rulemanager window.

Task profiles

A workflow consists of a series of tasks that defines a way to process documents. Datacap applications can include multiple workflows. This class focuses on single workflow applications.

A standard workflow for processing documents from Datacap Desktop takes a batch of documents through each of the processing steps.

A job consists of one or more tasks. To process a batch of documents, you must run the batch through each task in the selected job. Some tasks (for example, Export) run without operator intervention, but others (for example, Verify) require an operator.

The job type that you select determines the tasks in the workflow. You can see the tasks that are associated with each job type by looking in the Workflow pane on the Datacap Studio Zones view. A typical workflow for a job includes five tasks:

- VScan
- PageID
- Profiler
- Verify
- Export

Each task is linked to a task profile and the task profile determines the order that tasks are processed.

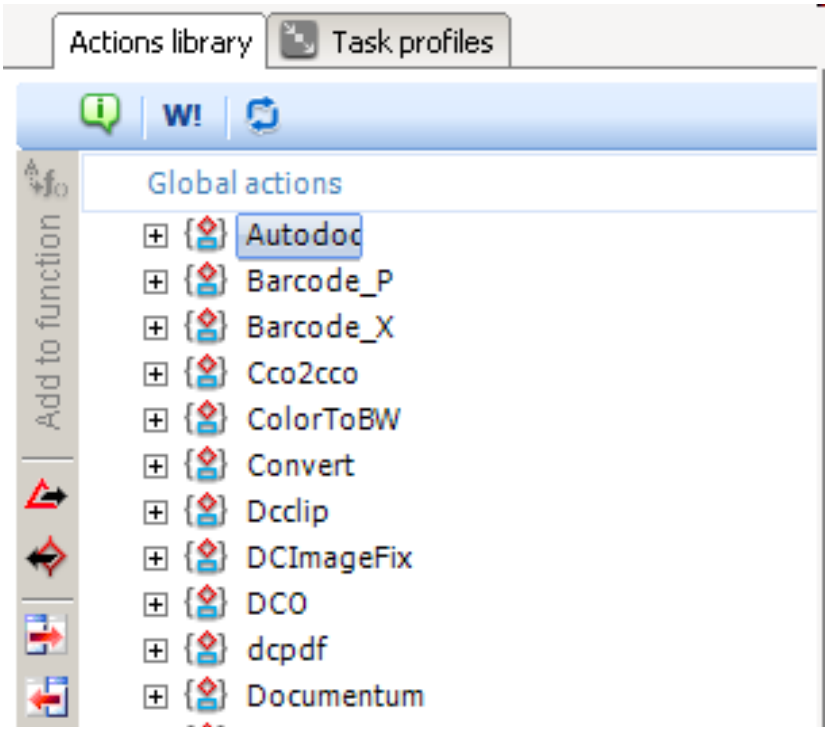
Add Rulesets to the Task profile.

Any new task that is defined must be added to a Task in the Task profiles for it to be included in the workflow for a batch. To add a ruleset to a task in the Task profile, you must select the ruleset in the Rulesets pane. Next, you must select the task in the Task profile pane that you want to be added to the selected task profile. Then, click the ‘Add ruleset to profile’ link on the left of the Task profiles pane.

IBM Training

Actions Library

- Tasks and Rulesets



Datacap Studio Introduction

Figure 1-21. Actions Library

The image on the slide illustrates the Actions library pane of the Datacap Studio Rulemanager window.

Actions library

Provides access to the complete library of prebuilt actions. To get help on an action, select the action and click Information. Each item that is listed in the Action library pane is a group of actions of a particular type. Click the plus sign next to each group to expand and show all the actions in the group.

Add actions to rules.

When building rules in the Ruleset pane, actions from one or more Action Libraries can be added to the functions. One or more functions makes a rule. To add actions to a rule in the Ruleset pane, select the function in the Ruleset pane that must receive the action. Next, you select the action to add to the rule. Then, click the Add to function link on the left of the Action library pane.

IBM Training

Demonstrations

- Start Datacap Studio
- Explore the Datacap Studio Rulemanager:
  - Document hierarchy tab
  - Rulesets tab
  - Task profiles tab
  - Properties tab
  - Locking the items for editing
  - Action libraries tab

Datacap Studio Introduction

Figure 1-22. Demonstrations

If you are taking this course as a self-paced virtual class, return to the main course menu to play the pre-recorded demonstrations.

IBM Training

Exercise: Explore Datacap Studio - Rulemanager

Requires:
Course Exercise Guide
Student system

Datacap Studio Introduction

Figure 1-23. Exercise: Explore Datacap Studio - Rulemanager

IBM Training

Exercise objectives

- Explore the Datacap Studio - Rulemanager

Figure 1-24. Exercise objectives

Lesson 1.2. Datacap Studio – Zones

IBM Training

Datacap Studio – Zones

Figure 1-25. Datacap Studio – Zones

IBM Training

Lessons

- Datacap Studio – Rulemanager
- Datacap Studio – Zones

Figure 1-26. Lessons

IBM Training

Why is this lesson important to you?

- As a Application Developer, you build and deploy applications with the Datacap capture system.
- To do these tasks effectively, you must be familiar with the Datacap Studio interface.

Figure 1-27. Why is this lesson important to you?

IBM Training

Zones View

- Fingerprints – Displays the application's fingerprint library.
- Document hierarchy – Defines the structure of the documents you are processing and how each element within the structure is processed.
- Properties –You can manage the properties for the selected document hierarchy object. It also lets you specify recognition options for the selected object.
- Image View – Shows the selected fingerprint image and any recognition zones.

Figure 1-28. Zones View

The Zones View is where you create page fingerprints and set up recognition zones. The Zones view includes the following panes:

Fingerprints

Shows the application fingerprint library and lets you add fingerprints for new page types.

Document hierarchy

Defines the structure of the documents that you are processing and how each element within the structure is processed.

Properties

Shows the properties for the selected document hierarchy object. If the document hierarchy is locked for editing, you can also modify existing properties.

The Properties pane also lets you specify recognition options for the selected object. Datacap Capture supports multiple recognition engines. Panes for ICR/C, BAR/P, and OCR/S are shown by default. You can access other panes by right-clicking within the Properties panel and selecting Show panes.

Image view

Shows the selected fingerprint image and any recognition zones. This view is where you draw new recognition zones. If you created the fingerprints with full page recognition, the Text pane at the bottom lets you view the recognition results.

IBM Training

Fingerprints View

- Displays the application's fingerprints library.
  - The fingerprint library is managed from here.
- Fingerprints classes and fingerprints can be added and deleted from this view.
  - This is where anything having to do with fingerprint management is done.

Figure 1-29. Fingerprints View

IBM Training

Document Hierarchy View

- The document hierarchy describes the structure of the documents that your application is designed to process.
  - The levels within the hierarchy are batch, document, page, and field.
- It also displays how each element in the structure is related to the other elements and how each object is processed.

Figure 1-30. Document Hierarchy View

IBM Training

Properties View

- Displays the properties for the selected document hierarchy object.
  - If the document hierarchy is locked for editing, existing properties can be modified.
  - In the properties panel, recognition options for the selected object can be specified.
- Since Datacap supports multiple recognition engines, the properties panel can display ICR/C, BAR/P, and OCR/S panels by default.
  - Other panels can be accessed by right-clicking within the Properties panel and selecting Show tabs.

Figure 1-31. Properties View

Image View

- Displays the selected fingerprint image and any recognition zones.
  - New zones can be drawn or existing zones corrected or changed in the image view panel.
- If fingerprints were created by using full page recognition, the recognition results can be seen in the Text tab.

Figure 1-32. Image View

IBM Training

Demonstrations

- Datacap Studio - Zones Tab

Figure 1-33. Demonstrations

If you are taking this course as a self-paced virtual class, return to the main course menu to play the pre-recorded demonstrations.

IBM Training

Exercise: Datacap Studio Zones

Requires:
Course Exercise Guide
Student system

Datacap Studio Introduction

Figure 1-34. Exercise: Datacap Studio Zones

IBM Training

Exercise objectives

- Explore the Datacap Studio – Zones view

Figure 1-35. Exercise objectives

IBM Training

Unit summary

- After completing this unit, you should be able to:
- Describe the components in Datacap Studio for application development process
- Use the Datacap Studio

Figure 1-36. Unit summary

Unit 2. Basic Learning Application

Estimated time

07:00

Overview

This unit introduces you to the process for building a basic Datacap learning application. It then walks you through the steps of refining the default image capture process as you make changes at each step of the process.

How you will check your progress

- Successfully complete the activities in the Student Exercises book.

References

IBM Knowledge Center

http://www.ibm.com/support/knowledgecenter/SSZRWV_9.0.1/com.ibm.datacaptoc.doc/datacap_9.0.1.htm

IBM Training

Unit Objectives

- After completing this unit, you should be able to:
- Create a basic learning application with the Datacap Application wizard
- Apply image enhancement techniques and field recognition potential
- Define a document structure
- Create and expand Locate rules
- Create and expand Validate rules
- Export the scanned documents to an IBM FileNet repository

Figure 2-1. Unit Objectives

Lesson 2.1. Create a Learning Template application

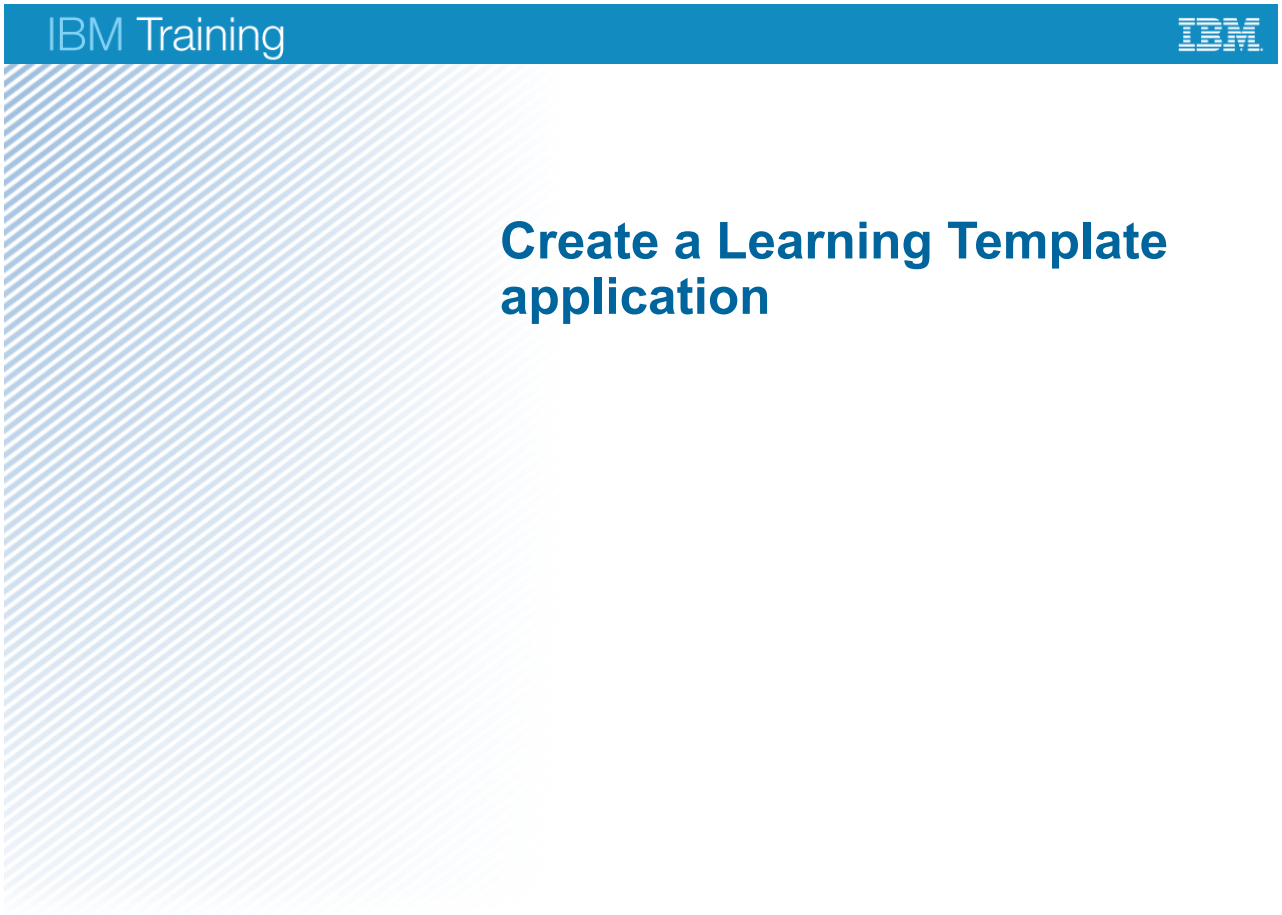


Figure 2-2. Create a Learning Template application

IBM Training

Lessons

- Create a Learning Template application
- Configure a Learning Template application
- Create locate rules
- Create validate rules
- Expand the locate rules
- Expand the validate rules
- Export to an IBM FileNet repository

Figure 2-3. Lessons

IBM Training

Why is this lesson important to you?

- As a Datacap business analyst, you build and deploy applications with the Datacap Capture system and communicate solution details to the solution architect, administrator, and users.
- This lesson gives you the ability to decide when to use a Learning template rather than a Form template when you create your application. It also gives you the ability to create the learning application by using the application wizard and to do the basic application configuration.

Figure 2-4. Why is this lesson important to you?

IBM Training

Create a Learning Template-based application

- Use the Application Wizard in FastDoc to create an application by using the Learning Template.
  - Name the application ‘Expense’
  - Select the Learning Template when prompted to select the Application Template
  - Verify that no errors or warnings are encountered

Figure 2-5. Create a Learning Template-based application

Create new Application with Application Wizard

The Application Wizard can also be started from Datacap Studio.

The Application Wizard Gives you a head start on application development by generating a basic application framework.

- Copy an application – What do you get with this option?
  - A copy of a completely functioning application as a starting point.
  - You can modify and customize it to meet the functional requirements for your new project.
- Create an RRS application – What do you get with this option?
  - An option of selecting one of two built-in application templates as a starting point for your new application.
  - A complete application folder structure.
  - A skeleton Document hierarchy (DCO), which you can expand to provide the document structure for your planned document batches.
  - A functional set of built-in compiled rulesets.
  - A set of workflow task profiles, preconfigured to run data capture tasks.

- A custom mapping configuration to map rules from the rulesets to the DCO objects.

Note:

1. You can create and save your own custom templates to use for future projects.
2. When creating new DCO objects, you can configure them to inherit rulesets from existing objects of the same type.

IBM Training

Characteristics of a Learning Template application

- A Learning Template-based application is inherently different from a Form-based application.
- It processes documents entirely differently
  - It has both different setup and processing
- It is called a Learning Template because the application learns how and when to create fingerprints and does it automatically.
  - You might add a few fingerprints when developing your application, but after that, leave it to the built-in behavior of the Learning Template.
  - It is not recommended to explicitly configure fingerprints

Basic Learning Application

Figure 2-6. Characteristics of a Learning Template application

Learning Template considerations

The first thing that you do when developing a new Datacap application is to decide which of the two built-in templates you should use. It isn't a difficult decision! A forms-based document type is one on which you know the types of data that you want to capture and where that data is on each image. Your documents are structured and consistent.

IBM Training

Characteristics of a Learning Template DocType

- Characteristics of a Learning Template document type.
  - The images are unstructured.
  - The types of data to capture are known.
  - The location of that data isn't known because it is different on each image.
    - For example, if you want to capture the date, amount and tax from different hotels, the receipt images for each hotel type are unique.
    - The data location to capture differs for each hotel receipt image, so fingerprints are not reliable at the start of building the application.

Figure 2-7. Characteristics of a Learning Template DocType

Document Type Considerations

Documents that are appropriate for a Learning Template-based application are unstructured. Documents that are made up of images which contain data that you want to capture but that data location isn't consistent. The location of the data might be different on every image.

IBM Training

Learning Template configuration

- The learning template uses generic document types, page types, and field types that are provided with the template.
  - It is important that you leave everything as it is except for the field objects that you define.
- There are four pages in the generic document object.
  - Main_Page
  - Trailing_Page
  - Attachment_Separator
  - Attachment
- It is important that you never change the Main_Page name.
  - Other page names can be changed although it's probably best not to do so!

Figure 2-8. Learning Template configuration

Learning Template Configuration.

- The learning template configuration uses generic document and page types that are provided with the template.
- The template contains four pages as the content of the generic document object:
  - Main_Page
  - Trailing_Page
  - Attachment_Separator
  - Attachment
- It is critical that you do not change those application objects – especially the Main_Page!
- All content must be put on the Main_Page

IBM Training

Main_Page

- Because you are developing a Learning Template application, you leave the name of the first page Main_Page.
- Main_Page represents two actual page types in the application you develop.
  - Rental_Agreement page
  - Air_Receipt page
- The only application objects that have custom names are fields.

Figure 2-9. Main_Page

Main_Page and your application

During the processing of your documents, the Main_Page page object will represent either the Rental_Agreement page type or the Air_Receipt page type. However, your application process only the pages of your document that uses the Main_Page as the representation of one of those application page types.

IBM Training

Document integrity Values

- It is important, when you define a new page of a document, that you define its document integrity values.
- Each page is part of a document object.
- Document integrity answers the questions:
  - Is there a minimum number of this page type that must be in the document? A maximum number? Must it always be the first page?
  - Can there be a variable number of this particular page type?
- You must configure the parameters that specify answers to those questions.

Figure 2-10. Document integrity Values

Document Integrity Overview

The Document Integrity ruleset that is generated by the Application Wizard is configured to identify document integrity problems and send the batch to the FixUp task if required.

IBM Training

Document integrity

- It is critical to define document integrity correctly.
  - It is how document integrity problems are found so that a document with integrity problems can be sent to Fixup.
- The three parameters that define document integrity for each page of a document are Minimum, Maximum, and Order.
- In the application, you are developing, each document has one, and only one, Rental_Agreement OR Air_Receipt page.
  - It is always the first page if some of your documents are multi-page documents.
  - Therefore, all three parameters Minimum, Maximum and Order each has a value of one.
  - These three parameters are on the Page Settings tab on the FastDoc 'Configure documents, pages, and fields' view.

Basic Learning Application

Figure 2-11. Document integrity

Defining Document Integrity

Three Document Integrity parameters that must be defined:

Minimum – the minimum number of a particular page type in a document type

Maximum – the maximum number of a particular page in a document type

Order – where this page must be in the document object

In the Learning Template labs, you set Minimum, Maximum, and Order to have the value of 1. In other words, there must be 1 and only 1, Rental_Agreement or Air_Receipt page, and that page is always the first page in the document.

IBM Training

Learning Template fields (1)

- Fields are the only application content that isn't entirely generic in a Learning Template-based application.
  - It is important that you leave batch structure objects rules and rule associations as they are, except for the field objects that you define.
- You define four generic fields on the Main Page
  - Name
  - Date
  - Total
  - Reference_Number

Basic Learning Application

Figure 2-12. Learning Template fields (1)

Generic fields

- Generic fields represent fields on one or more page types that aren't consistently in a particular location on a page.
- The fields are at different locations on different page types.
- The actual backend 'name' of the field on either page is not the generic field name at the point where you commit your document to that backend.
- It is critical to know the relationship between the processing name of a field object and the field's name on the backend to which you commit each document.

IBM Training

Learning Template fields (2)

- The four generic fields you define on the Main Page represent the following fields on the two page types

GENERIC FIELD NAME	Rental_Agreement Page	Air_Receipt Page
Name	Customer Name	Passenger Information Psgr
Date	Pickup Date/Time:	Issue Date:
Total	Total Charges	Total:
Reference_Number	Rental Agreement Number	TICKET #:

Figure 2-13. Learning Template fields (2)

Generic fields and how they relate to fields on the page of a document in a particular document class.

You have four generic field names that represent actual fields with different names in two different document classes.

This chart shows the relationship between the generic field name, “Name”, and the actual field name on two different page types.

For example, Name represents “Customer Name” on the Rental_Agreement page type and represents “Passenger Information” or “Psgr” on the Air_Receipt page type.

IBM Training

Exercise: Create a Learning Template application

Basic Learning Application

Figure 2-14. Exercise: Create a Learning Template application

IBM Training

Exercise objectives

- Create a Learning Template application

Figure 2-15. Exercise objectives

Lesson 2.2. Configure a Learning Template application

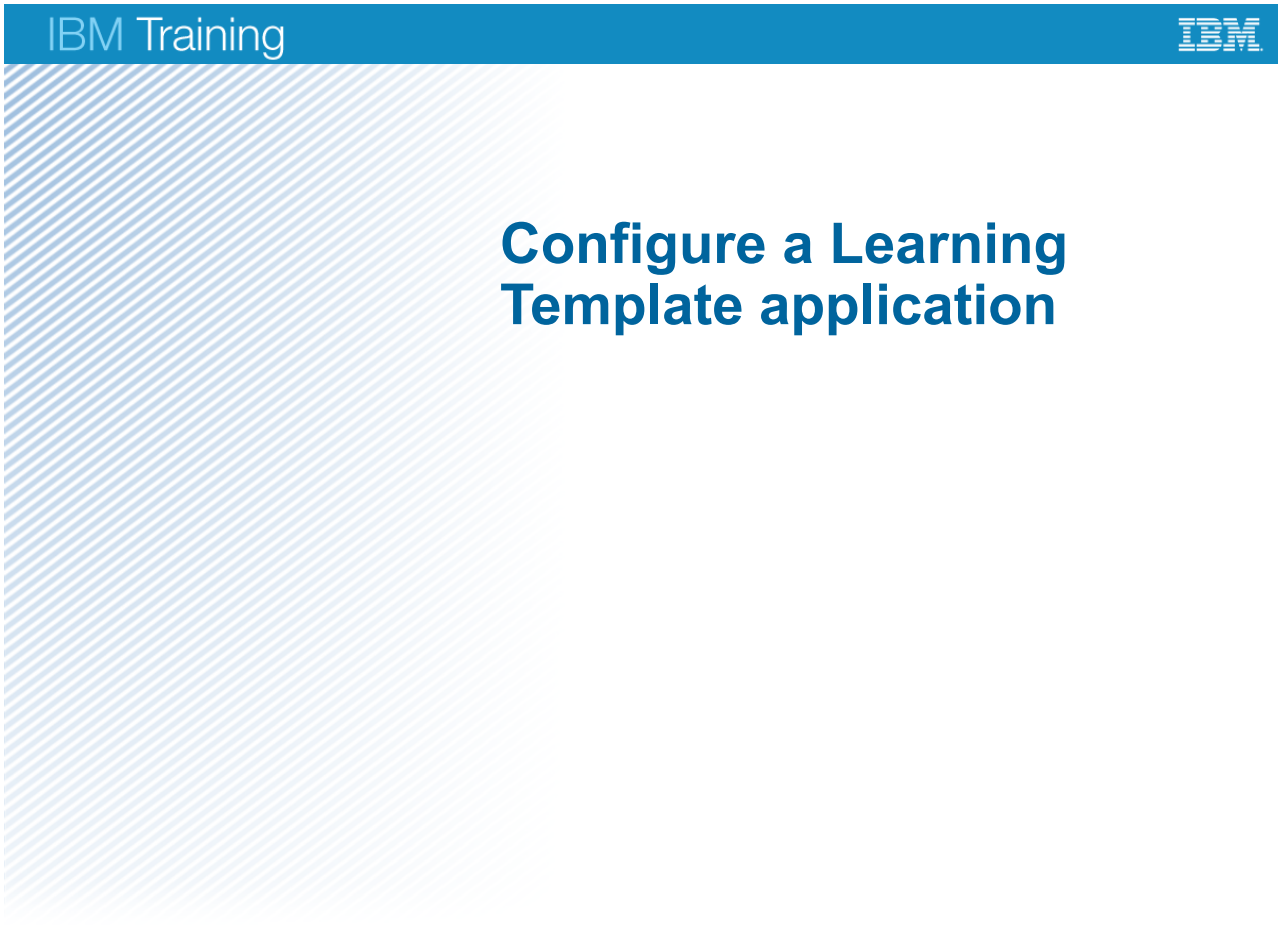


Figure 2-16. Configure a Learning Template application

IBM Training

Lessons

- Create a Learning Template application
- Configure a Learning Template application
- Create locate rules
- Create validate rules
- Expand the locate rules
- Expand the validate rules
- Export to an IBM FileNet repository

Figure 2-17. Lessons

IBM Training

Why is this lesson important to you?

- As a Datacap Solution/Application Builder, you build and deploy applications with the Datacap capture system and communicate solution details to the solution architect, administrator, and business analysts.
- When creating new applications, it is good practice to use the provided templates as a starting point. The Learning Template is used when processing images that differ in format but contain the same content to extract for indexing.

Figure 2-18. Why is this lesson important to you?

IBM Training

The Learning Template and fingerprints

- There is no need to set up fingerprints with this template.
  - If you do need to setup fingerprints, do so at the start of building the application by clicking and keying, for example:
  - You should not be generating fingerprints beyond the initial application creation and early configuration!
- Fingerprints are automatically generated as a user inputs data for the various document types and versions.

Figure 2-19. The Learning Template and fingerprints

IBM Training

The Learning Template workflow

- The Learning Template application sets up a workflow.
  - Rules like Locate rules are used to teach Datacap about the different page type formats as they are encountered.
  - Fingerprints are automatically generated when page types are determined.
- For images where data isn't found, the verifier is prompted to click the image and identify where that data is located.
  - This Click'N'Key process populates the data into the data set so that the Learning application can automatically find the data the next time that type of image is encountered.

Figure 2-20. The Learning Template workflow

The following procedure is the workflow process that the **Learning Template** uses for multi-page variable sized documents that use the **DemoMultiFormat** job.

1. Start FastDoc in Datacap Server mode. 2.
2. Log in to the **Learning** application.
3. Click the VScan shortcut. This task runs a disk scan on the image files that are stored in the C:\Datacap*appname*\images\input folder.
   a. Select the DemoMultiFormat job.
   b. Runs the **Import**, **Convert Images**, and **DeleteSourceImages** rulesets. FastDoc returns to the **Shortcuts** screen.
4. Click the Profiler shortcut. This task sets up the form by determining where the data that you want to process is on the image.
   a. Runs the Managed Rotation ruleset to enable rotation of the image if required.
   b. Runs the Image Enhancement ruleset to clean up the imperfections in the image. Select Auto Rotate to enable the Managed Rotation ruleset.

c. Runs the PageID ruleset to identify the path on which the document came in and creates the document structure. If the path is the same as another document, assumes that it is the same document type.

d. Puts all of the data on the Main_Page page type and treats every image as its own document. When the ruleset gets to a Separator sheet, it treats the next page as Main_Page. The pages after that are treated as a Trailing_Pages until the next Separator sheet is encountered.

   - Separator sheets are the method for grouping multiple pages into a document when using the Learning Template.

e. Runs the Recognize ruleset to identify the page types as Main_Page and Trailing_Page.

f. Inserts a Doc_Separator page in front of the Trailing_Pages. .

g. Inserts an Attactment_Separator page in front of the attachments. Recognition is not run on the attachments.

h. Runs the Fingerprint ruleset to build a fingerprint with a fingerprint ID for each document type.

i. Runs the Locate ruleset to search CCO for data, such as Social Security Number, Invoice Number, Date.

j. Runs the Lookup ruleset to take the fingerprint ID and find the fingerprint class for it.

k. Runs the Validate ruleset on the fields to validate the data in every field and make corrections where needed.

l. Runs the Routing ruleset to route the forms that passed validation to the export task.

5. Click the Verify shortcut.

   a. Runs the Validate ruleset on the form to remember the field information and automatically run it the next time that it is recognized.

6. Click the Export shortcut to run the ruleset that you configured for this job.

   a. Runs the SetStatuses ruleset to set up the exporting status of the document.

   b. Runs the PreExport ruleset to use the specified export settings for the document.

   c. Runs the Export ruleset to write the file to a specified location on the disk.

   d. Runs the Export to FileNet Content Manager ruleset to store the file in the IBM® FileNet® Content Manager repository.

   e. Runs the Export to IBM Content Manager ruleset to store the file in the IBM Content Manager repository.

   f. Runs the ProcessExceptions ruleset to handle exceptions to regular processing. For example, what to do with rescans and unwanted documents.

IBM Training

More about the Learning Template workflow

- There is no separate PageID task.
  - It is part of profiler because the assumption is made that every page read will be identified.
  - Since that assumption is made, different techniques are used for identifying documents.
- The document structure in the Learning Template approach is determined one of two ways.
  - Assume when using DemoSingleTIFFs that every page is a new document unless there is a doc separator in front of it.
    - Simulates documents that would come in at scan time
    - When documents are coming in from email or fax, there are no separator pages.
  - If there is a doc separator in front of it, it uses DemoMultiFormat to process variable sized documents that contain at least two pages.

Figure 2-21. More about the Learning Template workflow

IBM Training

More about the Learning Template workflow

- When images are pulled in, the Learning Template goes through what is called Managed Rotation
  - Therefore, it is important to turn off rotation when configuring image enhancement.
  - The Learning Template uses the Nuance document processing engine, which is powerful.
- As powerful as the Nuance engine is, if you have lots of images, the engine eventually crashes.
  - That situation is “normal” and expected
  - It recovers and continues processing
  - Nuance Instructions warn against running batches in an unmanaged fashion.
- At a minimum, you should always do deskew.

Figure 2-22. More about the Learning Template workflow

IBM Training

More about the Learning Template

- 'Recognize Pages and Fields' ruleset does full page recognition.
  - It should never need to be changed.
- 'Create Documents' is another U/I ruleset that you never need to change.
- Typically, in this type of application, only the main page is fingerprinted.
  - There are certainly exceptions, so if the characteristics of your images require that multiple pages be fingerprinted, by all means do so!
  - Use separator sheets between multi-page documents if your images require that.

Figure 2-23. More about the Learning Template

IBM Training

What about routing and export?

- Routing is set up in the Built-in application.
  - There are no branches or child jobs.
- Do not ever do a quick export, which bypasses the verify step.
  - Always verify every document.
  - One of the reasons is that two clients might be confused because both are using Quick Books and their forms are almost identical.
  - With Learning Template applications it is a good practice to show every Main_Page to a user as part of verification.

Figure 2-24. What about routing and export?

IBM Training

More things to know

- The Locate and Verify rulesets are built in to the Learning Template.
- You always have to define locate and validate rules:
  - It is a good practice to always validate every field you're capturing
  - At a minimum, specify:
    - If the field is required.
    - If it can be blank.
    - Whether it is overridable.
    - These options are easy to set in FastDoc.

Figure 2-25. More things to know

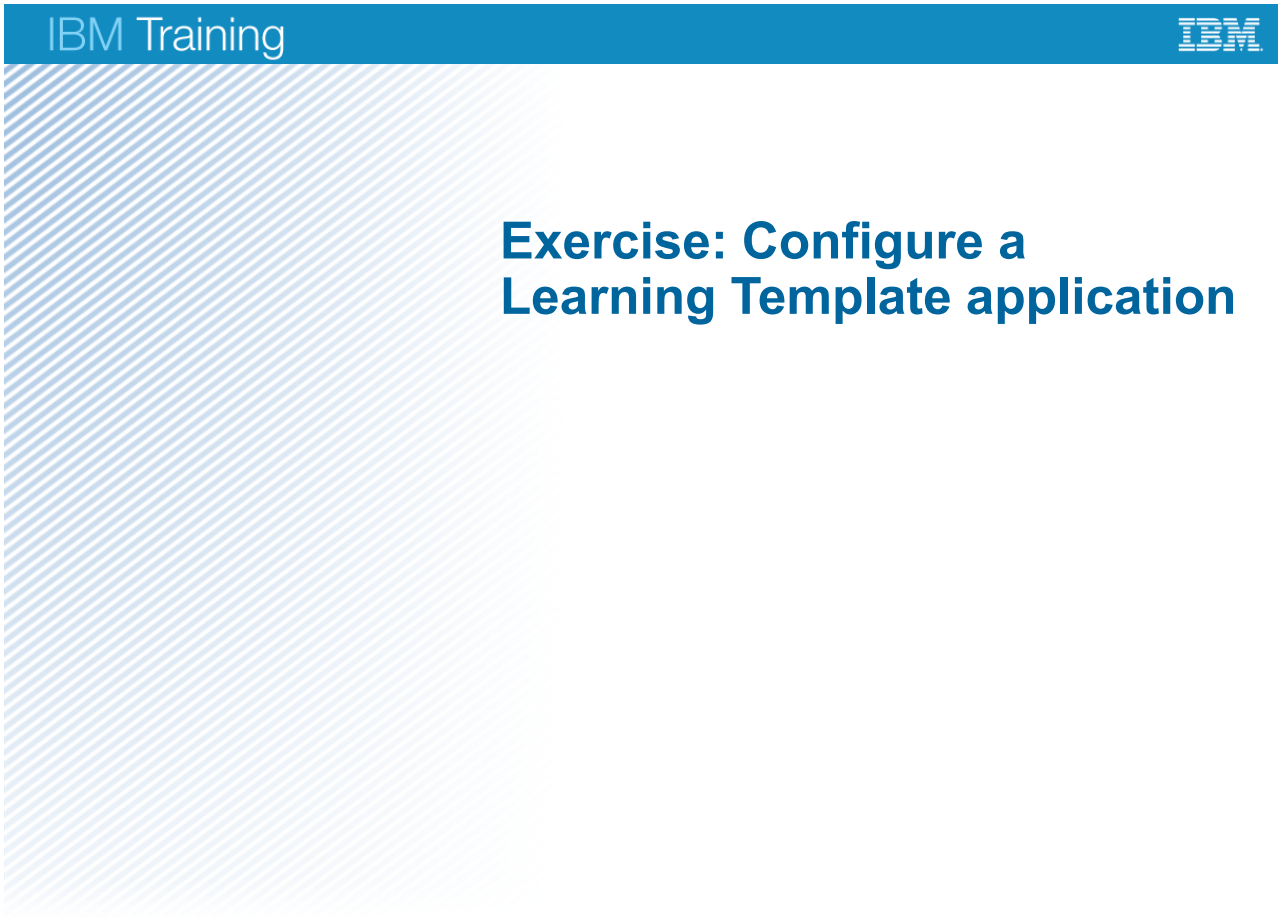


Figure 2-26. Exercise: Configure a Learning Template application

IBM Training

Exercise objectives

- Basic application configuration

Figure 2-27. Exercise objectives

Lesson 2.3. Create locate rules

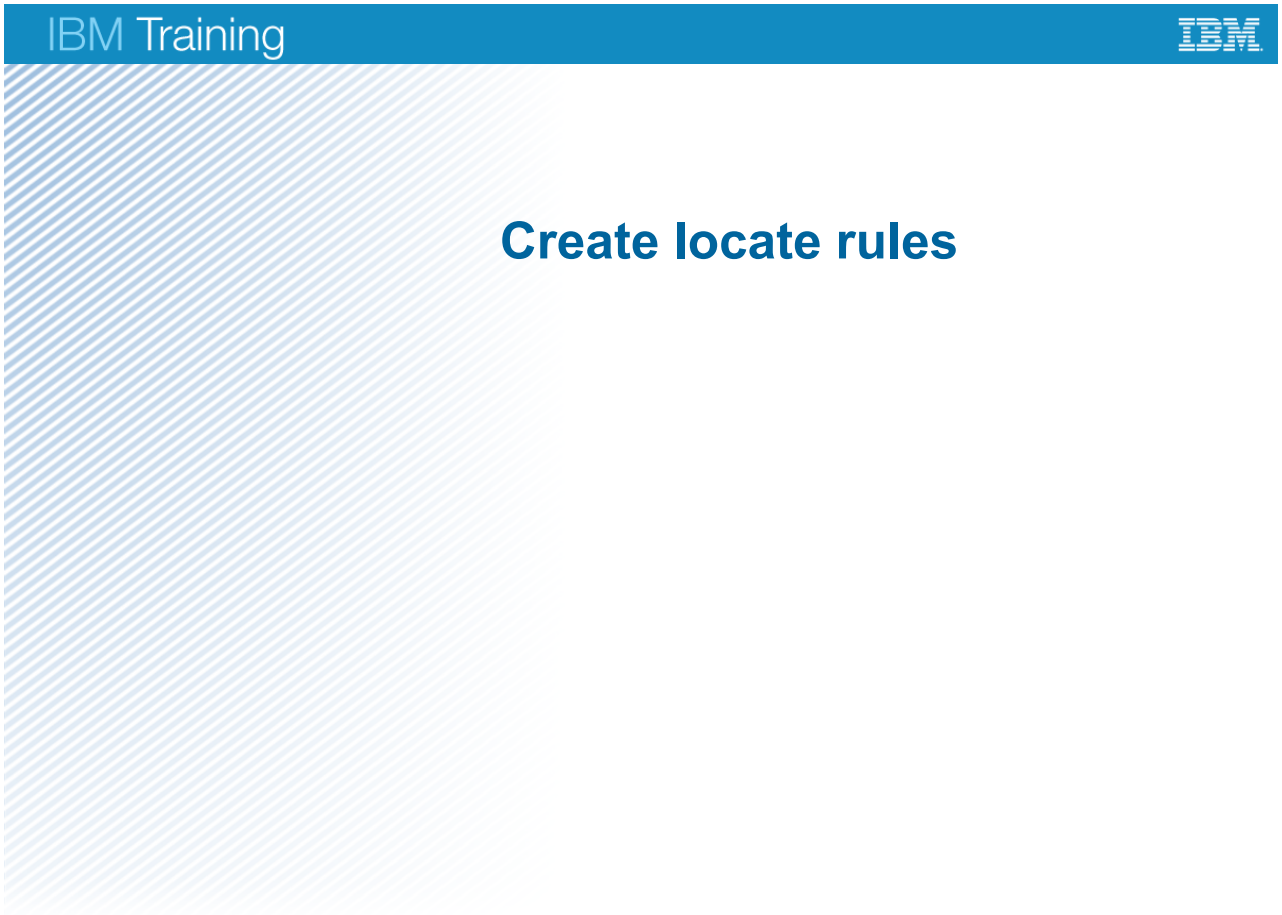


Figure 2-28. Create locate rules

IBM Training

Lessons

- Create a Learning Template application
- Configure a Learning Template application
- Create locate rules
- Create validate rules
- Expand the locate rules
- Expand the validate rules
- Export to an IBM FileNet repository

Figure 2-29. Lessons

IBM Training

Why is this lesson important to you?

- As a Datacap Solution/Application Builder, you build and deploy applications with the Datacap capture system and communicate solution details to the solution architect, administrator, and business analysts.
- When creating applications to process images that have inconsistent page layouts, you can save users time. You save them time by using locate rules to populate fields by using the click‘n’key method when data positions move from one page to the next.

Figure 2-30. Why is this lesson important to you?

IBM Training

The basic locate actions

- Locate ruleset
  - Use key-word-search to locate and extract data from a document.
  - Use zones that are marked during fingerprinting to directly locate and extract values from cco files created by full page zonal recognition.
- Actions to read and extract data from zones:
  - ReadZones: Loads the position information for the current object and its children from the document hierarchy.
  - PopulateZNField: Populates the current field value in the page data file with the recognition data from the CCO file that lies within the field zone boundaries.
- If fields never changed locations, then these two actions would correctly read and extract data from the fields on a page.
- These actions are included in the built-in Locate rule for the Learning Template.

Figure 2-31. The basic locate actions

Overview

A ruleset defines one or more rules that you can run on specific documents, pages, fields, or on the entire batch. You must associate a specific rule with one or more of the objects in the document hierarchy.

Key locate actions

The Locate ruleset is used to read and extract data from a document. Data can be directly read and extracted from the zones that are marked during fingerprinting the page or can be searched for using configured action and parameters.

Actions to read and extract data from zones:

ReadZones: Loads the position information for the current object and its child objects from the document hierarchy.

PopulateZNField: Populates the current field value in the page data file with the recognition data from the cco file that lies within the field zone boundaries.

Advanced Locate actions are presented later in this course.

Rule configuration and run explanation

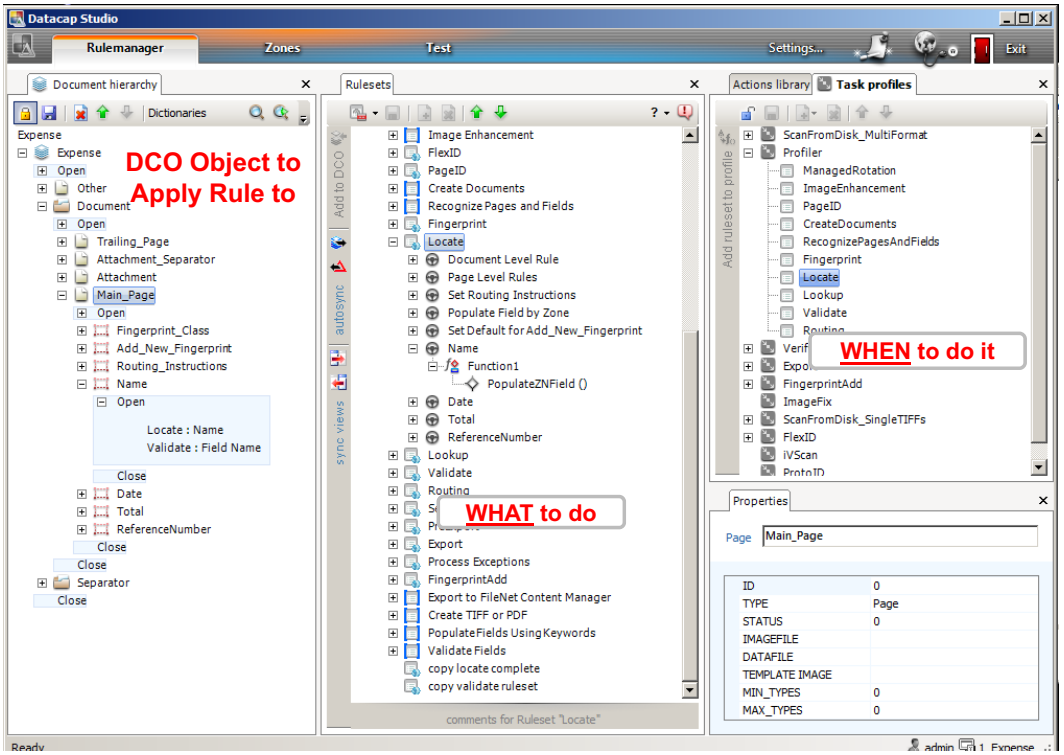


Figure 2-32. Rule configuration and run explanation

The image on the slide illustrates that:

- The Ruleset pane determines what must be done.
- The Task profiles determine when to do it.
- The document hierarchy pane determines which DCO objects to apply it to.

IBM Training

Locate actions (1)

Locate Actions	Description
FindKeyList	Locates the first (or next) occurrence of a word or phrase that matches one of the entries in a keyword file.
GoBelowWord	Moves down the specified number of lines from the previously found word or phrase.
GoLeftWord	Moves the specified number of words to the left of the previously found word or phrase.
GoRightWord	Moves the specified number of words to the right of the previously found word or phrase.
GroupWordsRIGHT	Groups words to the right of the previously found word if they are no more than the specified number of character widths apart.

Basic Learning Application

Figure 2-33. Locate actions (1)

Locate rule considerations

Locate rules can be used in place of fingerprinting to locate and extract image data. In this lesson, you develop locate rules for each image field and add them to the Locate ruleset. The Locate ruleset allows all of the fields to be identified on the image and for the data to be read. You must also associate the locate rules with the field definition in the DCO and define the Locate ruleset in the Task profile.

Locate Overview.

To recover data with locate rules, you need to create a rule for each field that is defined in the page definition on the Document hierarchy pane. Select fields that are not automatically populated with the two genetic rules that are already added to the Locate ruleset. Some of the advanced rules that can be used for this purpose are reviewed here. You are going to use the following Locate and Zone actions to find data.

Locate actions (2)

Locate Actions	Description
IsCurrency	Determines whether the value of the located word is a currency value (is numeric and includes a two-digit decimal amount).
IsDateValue	Determines whether the value of the located word is in one of the supported date formats.
MinLength	Determines whether the number of characters in the located word is less than or equal to the number specified (returns True if it is).
MaxLength	Determines whether the number of characters in the located word is greater than or equal to the number specified (returns True if it is).
RegExFind	Same as WordFind, but it supports regular expressions.

Figure 2-34. Locate actions (2)

IBM Training

Locate and zone actions

Locate Actions	Description
UpdateField	Updates the current field in the page data file with the value and position of the located word.
WordFind	Locates the first (or next) occurrence of the specified word or phrase on the current page.

Zone Actions	Description
PopulateZNField	Populates the current field value in the page data file with the recognition data from the cco file that lies within the field zone boundaries.

Figure 2-35. Locate and zone actions

Rental_Agreement page

- Four fields to recognize and extract.

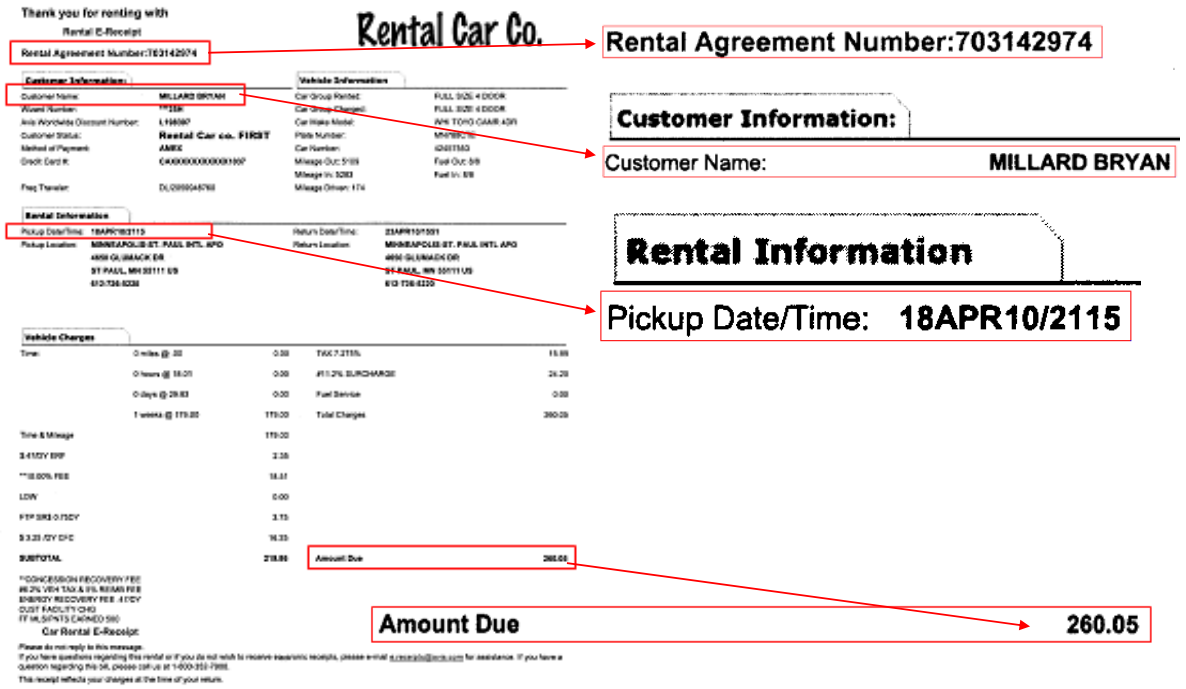


Figure 2-36. Rental_Agreement page

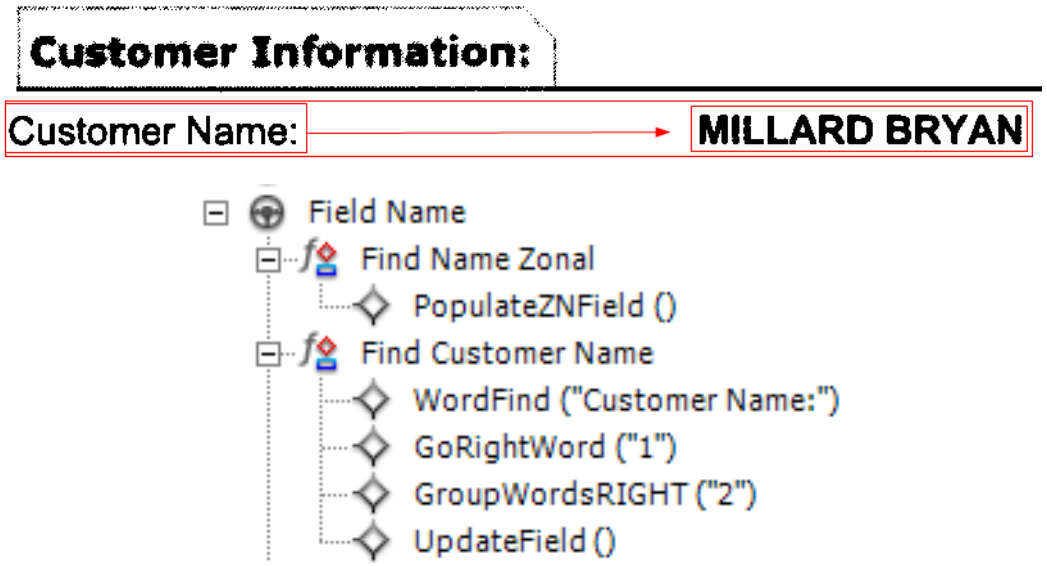
The images are:

- A view of the Rental Agreement page with the four fields marked that you create locate rules for in the lesson lab.
- An expanded view of each of the four areas so they are more readable.
  - Rental Agreement Number
  - Customer Name
  - PickupDate/Time
  - Amount Due

IBM Training

Locate rule for customer name field

- Add Locate rules for Customer Name field to extract data.
- Find Name Zonal
- Find Customer Name



Basic Learning Application

Figure 2-37. Locate rule for customer name field

The images are:

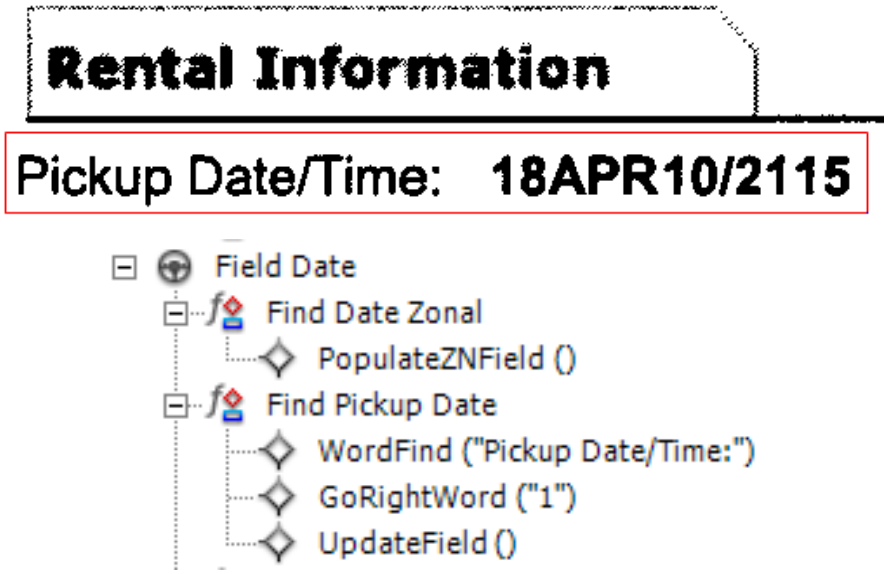
- The section of the rental agreement with the customer name.
- The Field Name rule for locating the field on the page.

1. Find Name Zonal
   a. Read the values from the zone (cco) file that is created by the full page zonal recognition action that is run at the page level.
2. Find Customer Name
   a. Search for the word “Customer Name:”.
   b. Move right one word.
   c. Read two words to the right, which are the passenger first and surnames.
   d. The action updates the current field in the page data file with the value read.

IBM Training

Locate rule for pickup date

- Add Locate rules to individual fields to extract data.
- Find Reference Number Zonal
- Find Rental Agreement



Basic Learning Application

Figure 2-38. Locate rule for pickup date

The images are:

- The section of the receipt with the pickup date and time .
- The Field Pickup Date rule for locating the field on the image.

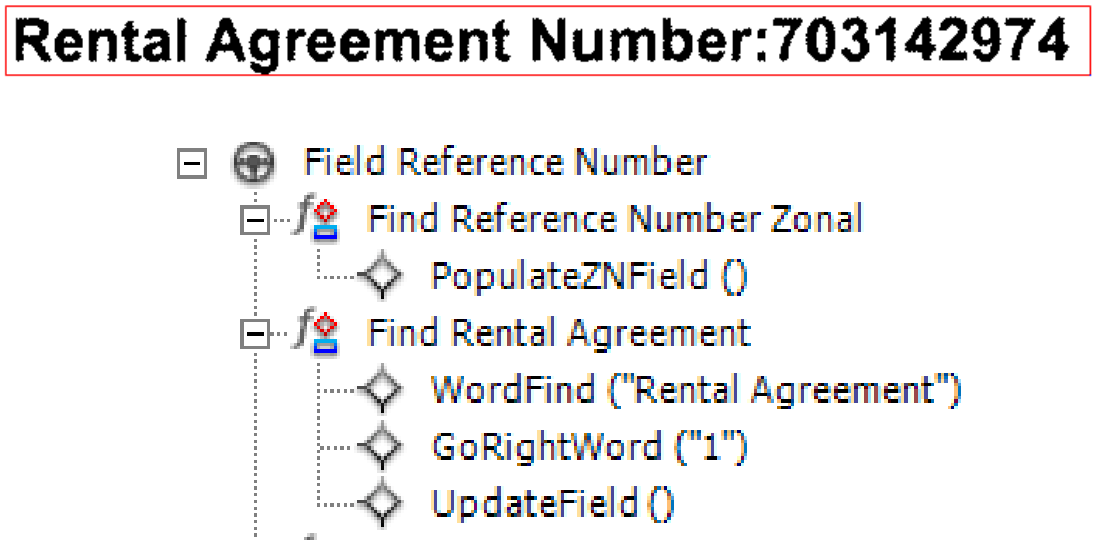
1. Find Date Zonal
   a. Read the values from the zone (cco) file that is created by the full page zonal recognition action that is run at the page level.
2. Find Rental Agreement
   a. Search for the text “Pickup Date/Time:”
   b. Read one word to the right, which is the whole Expiration date field.
   c. The action updates the current field in the page data file with the value read.

- When this rule is complete you, do not have a valid date yet because the time is still appended. In the following lesson when you create the validate rules, you implement actions that correct this problem.

IBM Training

Locate rule for agreement number field

- Add Locate rules for Agreement Number field to extract data.
- Find Reference Number Zonal
- Find Rental Agreement



Basic Learning Application

Figure 2-39. Locate rule for agreement number field

The images are:

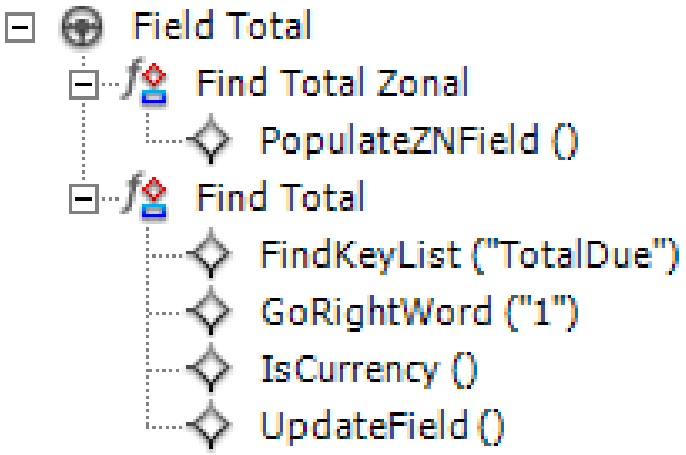
- The section of the page with the Rental Agreement Number.
- The Field Reference Number ruleset for locating the field on the image.

1. Find Rental Agreement Zonal
   a. Read the values from the zone (cco) file that is created by the full page zonal recognition action that is run at the page level.
2. Find Rental Agreement
   a. Search for the text “Rental Agreement”.
   b. Read one word to the right, which is Number:*9digit number*.
   c. The action updates the current field in the page data file with the value read.

IBM Training

Locate rule for total field

- Add Locate rules for total field to extract data.
- Find Total Zonal
- Find Total
- 



Basic Learning Application

Figure 2-40. Locate rule for total field

The images are:

- The section of the receipt with the Total amount.
- The Field Total rule for locating the field on the image.

1. Find Total Zonal
   a. Read the values from the zone (cco) file that is created by the full page zonal recognition action that is run at the page level.
2. Find Total
   a. Search for the words on the page that are a match any string that listed in the “**TotalDue**” key file. If a match occurs, move to the next action.
   b. Read one word to the right, which is the amount due for the rental agreement.
   c. Check that the value is in currency form.
   d. Update the current field in the page data file with the value and position.

- The TotalDue.Key file is in the C:\Datacap\<application name>\dco_<application name> folder.
- It contains a list of many strings that are possible search strings for a value that might represent the key word that matches the page total value.

- The sequence of processing is, for each string of text that is read from the page, search the .Key file for a match. If a match is found process the next action.

IBM Training

Demonstrations

- Create Locate rules in Datacap Studio

Figure 2-41. Demonstrations

If you are taking this course as a self-paced virtual class, return to the main course menu to play the pre-recorded demonstrations.

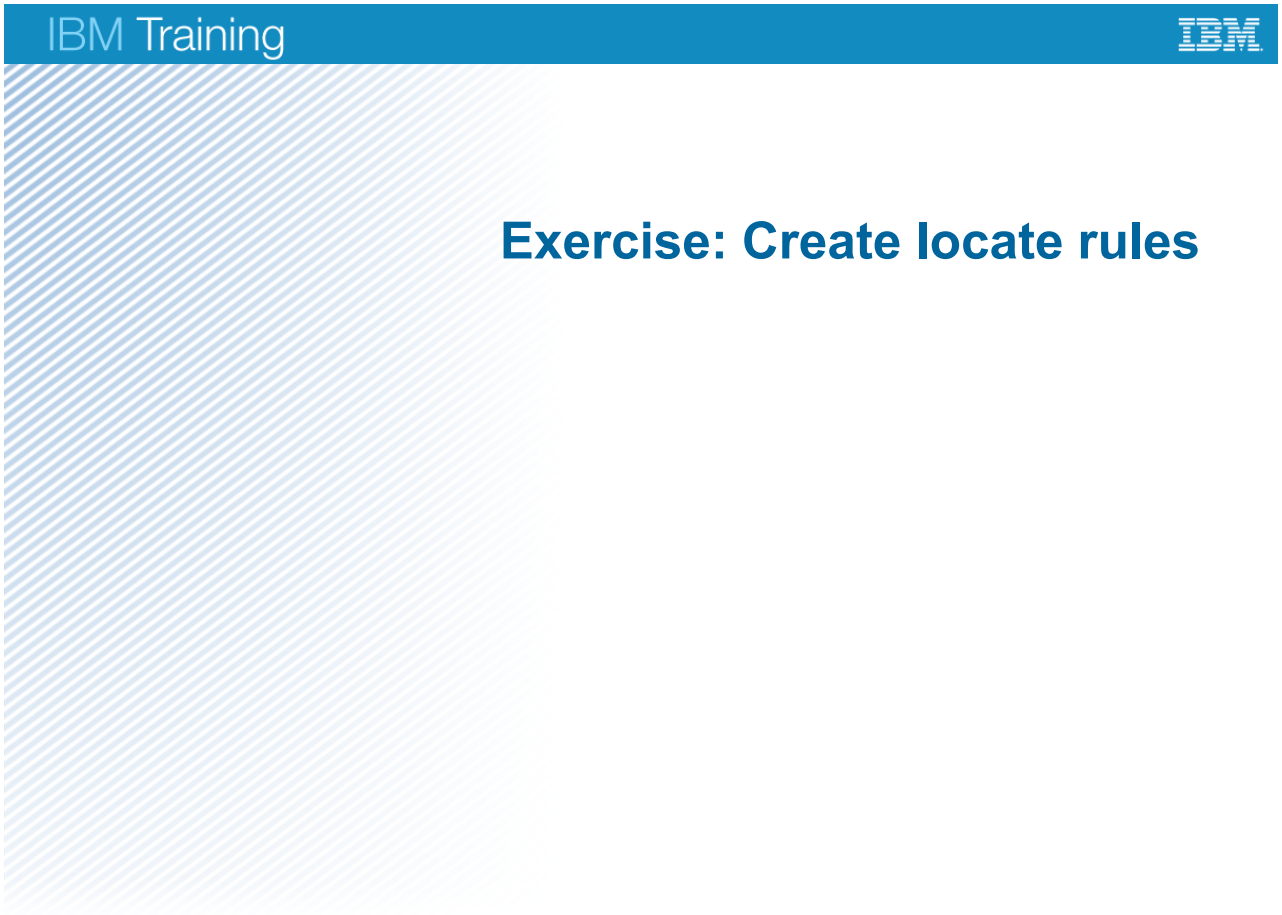


Figure 2-42. Exercise: Create locate rules

IBM Training

Exercise objectives

- Locate fields by using keyword search

Figure 2-43. Exercise objectives

Lesson 2.4. Create validate rules

Figure 2-44. Create validate rules

IBM Training

Lessons

- Create a Learning Template application
- Configure a Learning Template application
- Create locate rules
- Create validate rules
- Expand the locate rules
- Expand the validate rules
- Export to an IBM FileNet repository

Figure 2-45. Lessons

IBM Training

Why is this lesson important to you?

- Because you are working with the Learning Template, it is important to run validation rules on all the fields of the page you are configuring when you make your first pass through the application
- To successfully capture documents, you must be an expert at programmatically validating the captured data.

Figure 2-46. Why is this lesson important to you?

IBM Training

The validate ruleset

- The Learning Template has a built-in page level rule that is applied to the Main_Page page.
- The Validate > Page Level Rule runs Status_Preserve_OFF
  - Setting Status_Preserve_OFF on the Validate Page rule allows the validation actions to assign a “problem” status to an invalid or empty field value.
  - Reported error allow verifiers to correct potential value errors

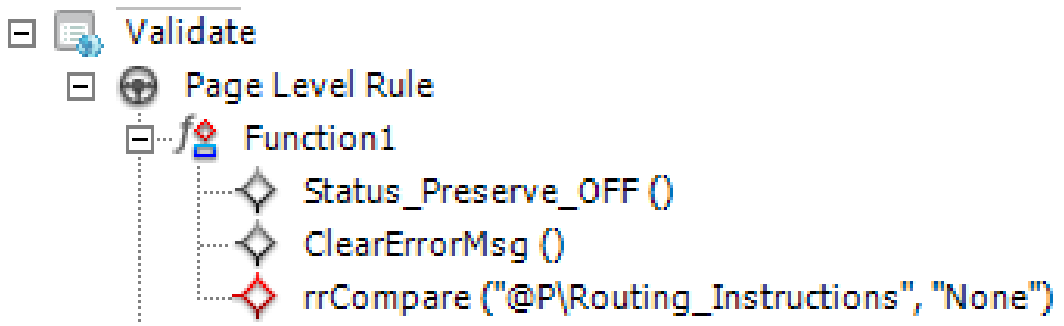


Figure 2-47. The validate ruleset

The image on the slide the ‘Page Level Rule’ uses the status_Preserve_OFF() action to ensure that field error are identified at the Verify step if errors occur.

Procedure to do simple validation and routing tasks.

Set Preserve OFF on Validate Page rule

In the default Validate Page rule, the action is set to Status_Preserve_OFF. You leave this rule unchanged. Having this rule present is important. It causes the batch to be routed to the Verify step if a field error is detected.

An Off condition of an object allows the actions of a Validate ruleset to assign a problem status to any field object with an invalid capture value. The Verify task Data Entry panel surrounds the value with a pink background, alerting the operator to the problem.

This error flagging occurs when a Validate field rule fails. The page status reflects the failure so that it shows at the Verification phase.

Add a field validation rule to a ruleset

Select the Validate ruleset in the Ruleset pane and lock the Ruleset structure for editing.

Right-click the validate Ruleset and click Add new rule. Change the rule name to identify the field that it operates on. For example, name the rule Field Name. The Field portion of the name identifies that the rule operates on a field node in the Document hierarchy. The Name portion of the rule name identifies that the field that is operated on by this rule is the name field.

IBM Training

Validate and export task relationship

- In addition to running validation as part of your first pass through the learning template application, it is important to also run export.
  - The location of each field on the Rental_Agreement page isn't saved until you run the export task.
- In the lab, you add a Field Validation Rule to each of the four fields in the application.
  - Some rules might require more than one function.

Figure 2-48. Validate and export task relationship

Overview

It is critical to run export to update the location of the fields on your images. The field locations are not saved in a new fingerprint until export is complete.

Sample Validation actions

- IsFieldLengthMax (5)
  - Are there more than 5 characters in the field?
- IsFieldLengthMin (1)
  - Is there at least one character in the field?
- AllowOnlyChars (“0123456789”)
  - Allow only numeric digits 0-9 in the field
- Calculate (“’SubTotal’ + ‘Shipping’+ ‘Tax’ = ‘Total’”)
  - Verify that the calculation based on field values is correct.
- IsFieldFilled ()
  - Is there any data in the field?
- IsFieldDate ()
  - Does the field contain a valid Date

Figure 2-49. Sample Validation actions

Assign an action to the Field Name rule

- Select function
- Select action
- Add to function
- Set action properties

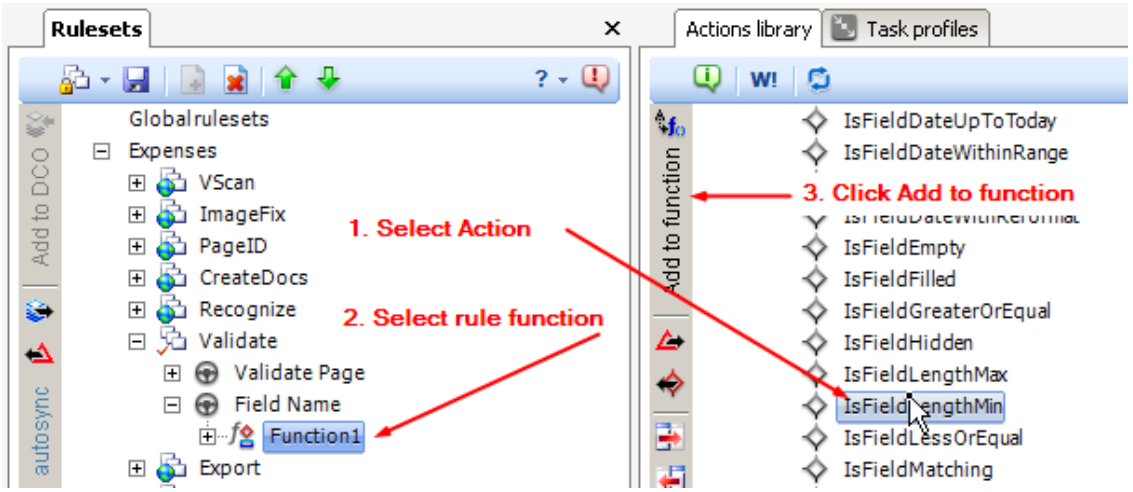


Figure 2-50. Assign an action to the Field Name rule

The Image on the slide shows a graphic representation off the method for assigning actions to functions in a rule.

Add an action that checks for a minimum field length of 5 in the name field.

Define a Validate rule for the Name field.

1. Select Function1 for the new Field Name rule that is selected in the Ruleset pane.
2. Select the IsFieldLengthMin action in the Validations Actions group of the Actions library pane.
3. Click the Add to function link.
4. Click the IsFieldLengthMin() action that you just added to the rule and set the Parameter to 5 on the properties pane.

IBM Training

Validations actions

Locate Actions	Description
AllowOnlyChars	Removes all characters that are not specified as allowed characters.
FailRuleSet	Causes the entire ruleset to fail.
IsFieldDate	Returns True if the current field value is a date.
IsFieldDateWithinRange	Returns True if the current field value is a date within the specified range.
IsFieldDateWithReformat	Returns True if the current field value is a date and reformats the value to the date format specified.
IsFieldFilled	Returns True if the current field value is not empty.

Basic Learning Application

Figure 2-51. Validations actions

Validate and route

Overview

When field data is located and extracted from the page image, it is necessary to validate the reliability of the data. This information is used to be sure that all fields are completed with an acceptable degree of accuracy. Validate rules are used at verify time to make sure that the operator does not enter invalid values.

Some of the actions that can be used for validation are reviewed here.

IBM Training

Validations actions

Validate Actions	Description
IsFieldLengthMax	Returns True if the length of the current field value is less than or equal to the length specified.
IsFieldLengthMin	Returns True if the length of the current field value is greater than or equal to the length specified.
LeftTruncate	Deletes characters from the end of the current field value until the length equals the length that is specified.

Figure 2-52. Validations actions

IBM Training

Mixing locate and validate actions

- It is acceptable to mix Locate and Validation actions in the same function.
- Locate actions operate on cco fingerprint values.
- Validation actions act on the field value.
- When mixing actions, the UpdateField() action must follow the locate actions; then the Validation actions follow UpdateField().
  - If this sequence is not followed, then the validation actions act on an uninitialized value.
- Reason for not mixing Locate and Validation actions.
  - Locate rule runs only in the Profiler task.
  - Validate rule runs in the Profiler and Verify task.
  - Any validation actions included in in the Locate rule are not run at the Verify step if you do run Validations.

Basic Learning Application

Figure 2-53. Mixing locate and validate actions

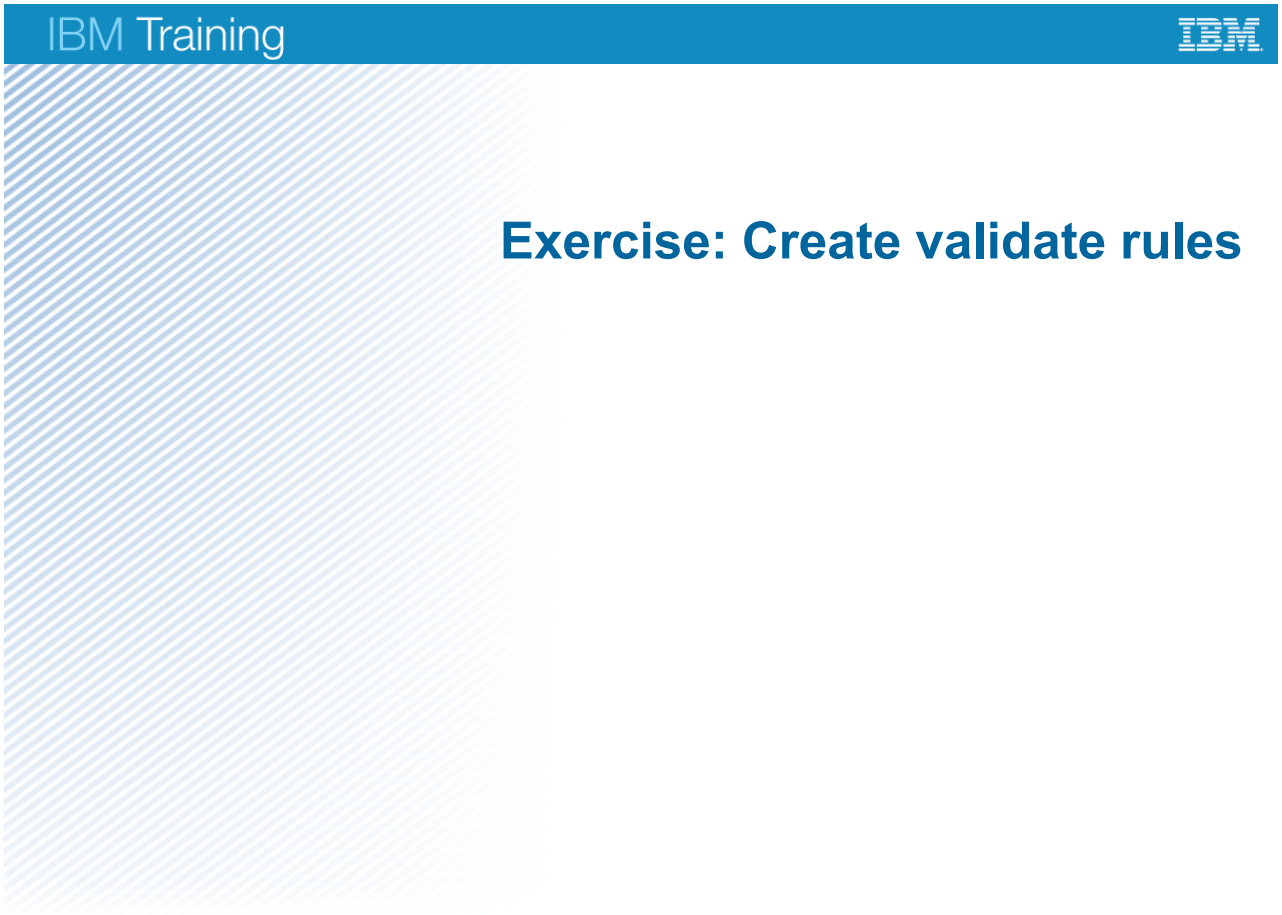


Figure 2-54. Exercise: Create validate rules

IBM Training

Exercise objectives

- Configure field validation rules

Figure 2-55. Exercise objectives

Lesson 2.5. Expand the locate rules

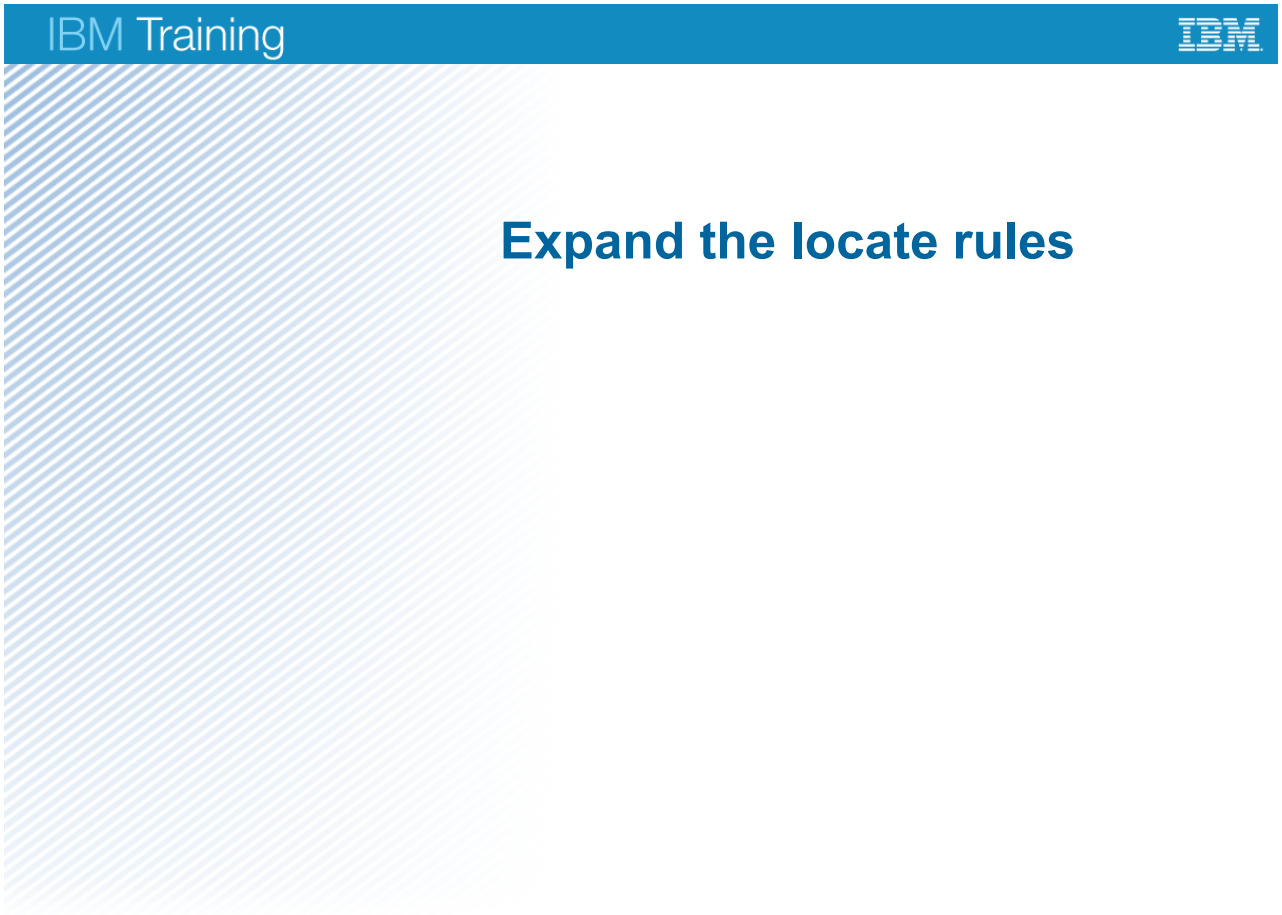


Figure 2-56. Expand the locate rules

IBM Training

Lessons

- Create a Learning Template application
- Configure a Learning Template application
- Create locate rules
- Create validate rules
- Expand the locate rules
- Expand the validate rules
- Export to an IBM FileNet repository

Figure 2-57. Lessons

IBM Training

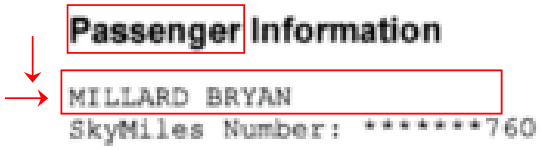
Why is this lesson important to you?

- As a Datacap Solution/Application Builder, you build and deploy applications with the Datacap capture system and communicate solution details to the solution architect, administrator, and business analysts.
- To build Datacap applications, you need to be an expert in the use of the locate actions. Locate actions are an alternative and sometimes more explicit method of identifying fields on a page.

Figure 2-58. Why is this lesson important to you?

Locate rule for passenger name field

- Add Locate rules for passenger name field to extract data.
  - Search for the word “Passenger”.
  - Step down one line.
  - Read two words to the right, which are the passenger first and last names.
  - The action updates the current field in the page data file with the value.



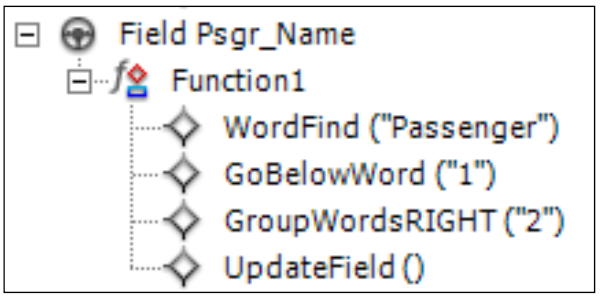


Figure 2-59. Locate rule for passenger name field

The images are:

- The section of the receipt with the passenger name.
- The Field Psng_Name rule for locating the field on the image.

IBM Training

Locate rule for the ticket number field

- Add Locate rules for ticket number field to extract data.
- Find Ticket #
  - Search for the text “TICKET #”.
  - Read one word to the right,

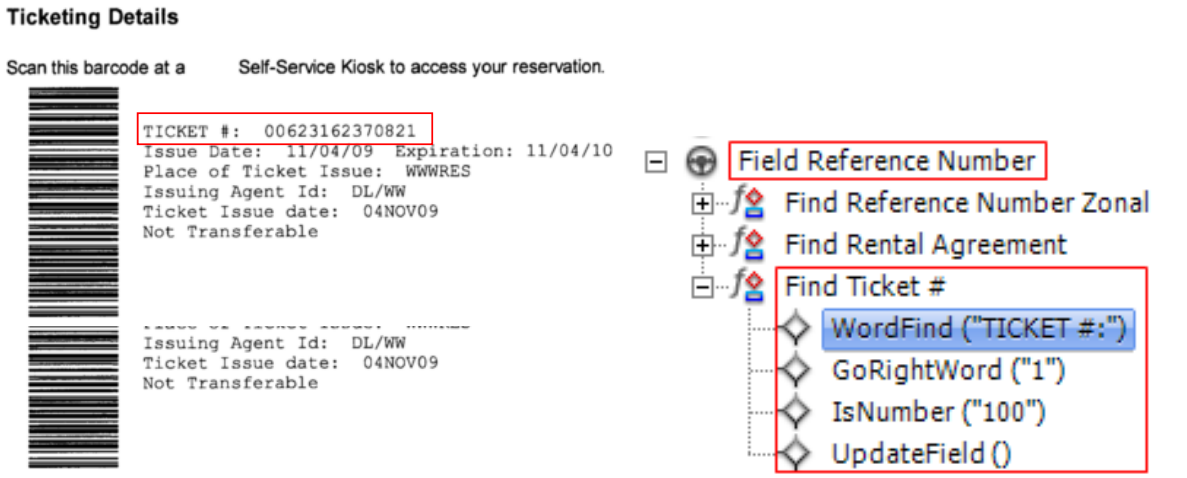


Figure 2-60. Locate rule for the ticket number field

The images are:

- The section of the receipt with the ticket number.
- The Field Ticket Number rule for locating the field on the image.

Find Ticket #

a. Search for the text “TICKET #”.

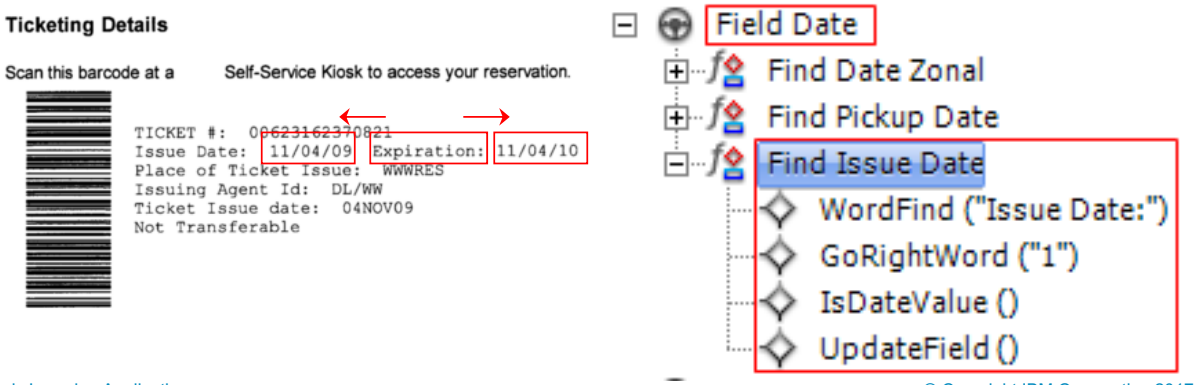
b. Read one word to the right, which is the 14-digit ticket number.

c. The action updates the current field in the page data file with the value.

IBM Training

Locate rule for issue dates

- Add Locate rules to individual fields to extract data.
  - Search for the text “Issue Date:”
  - Read one word to the right, which is the Issue Date field.
  - Verify that the value is a date.
  - The action updates the current field in the page data file with the value.



Basic Learning Application

Figure 2-61. Locate rule for issue dates

The images are:

- The section of the receipt with the Issue date and the Expiration date.
- The Field Issue Date rule for locating the field on the image.

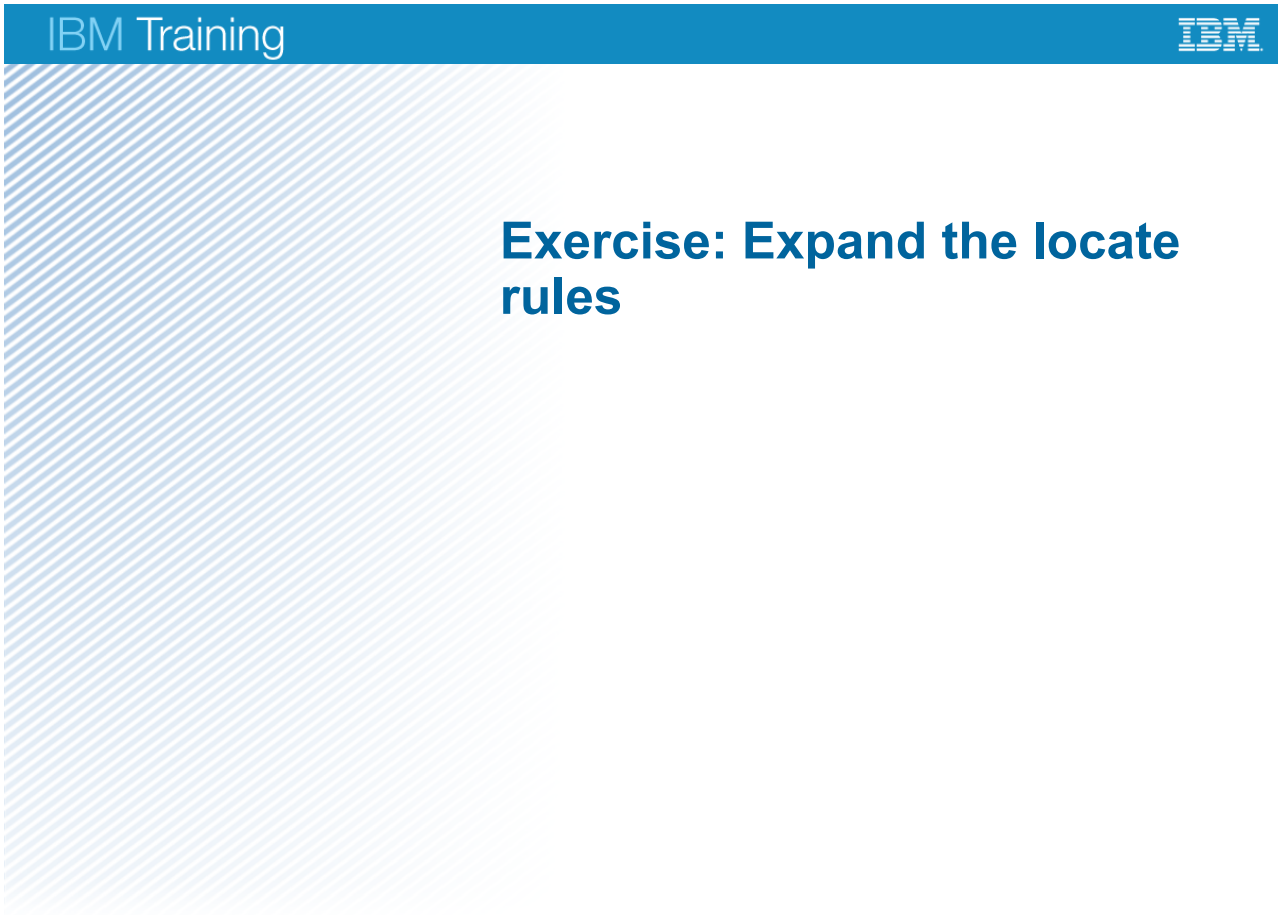


Figure 2-62. Exercise: Expand the locate rules

IBM Training

Exercise objectives

- Expand the Locate Ruleset for Air Receipt pages
- Create an Air_Receipt page fingerprint

Figure 2-63. Exercise objectives

Lesson 2.6. Expand the validate rules

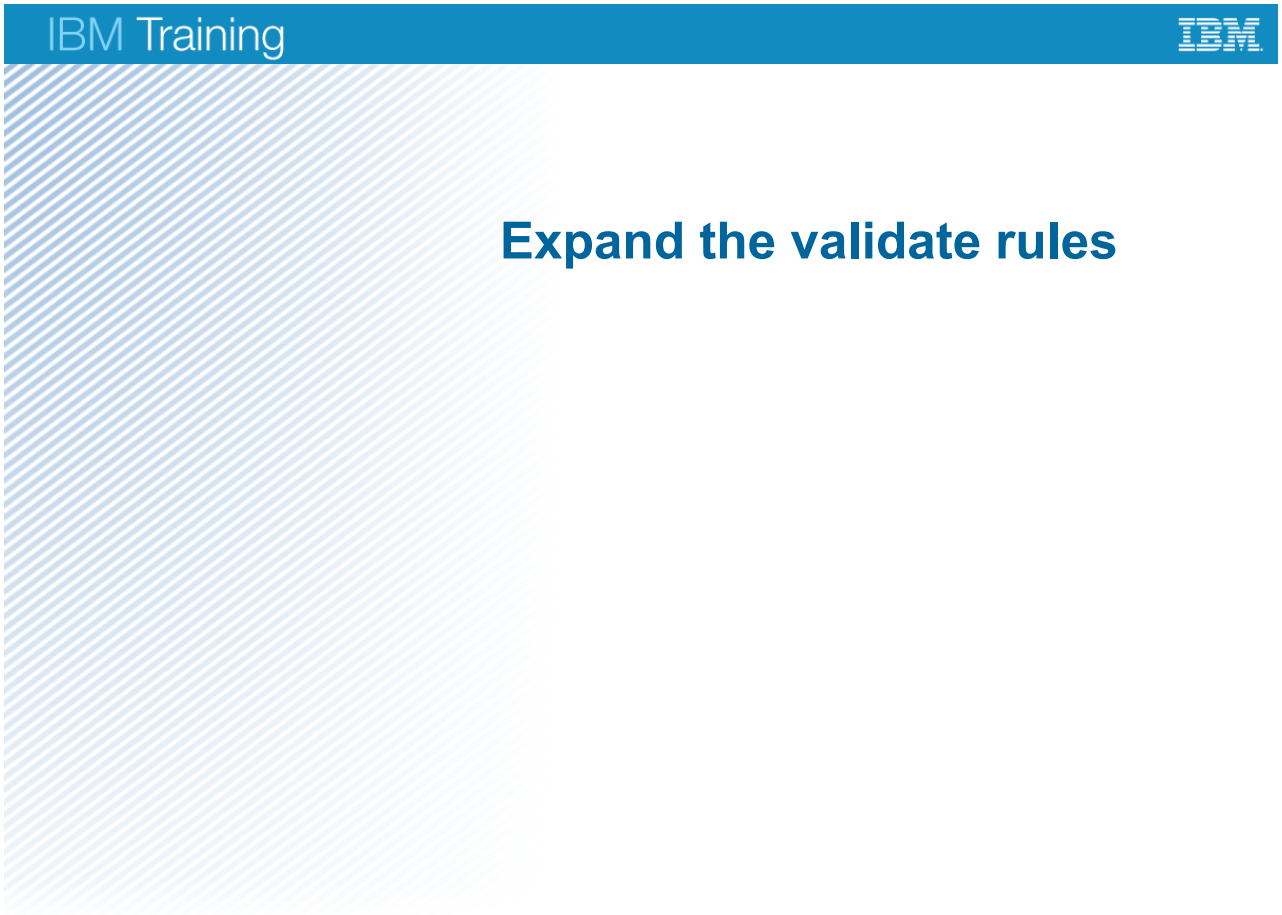


Figure 2-64. Expand the validate rules

IBM Training

Lessons

- Create a Learning Template application
- Configure a Learning Template application
- Create locate rules
- Create validate rules
- Expand the locate rules
- Expand the validate rules
- Export to an IBM FileNet repository

Figure 2-65. Lessons

IBM Training

Why is this lesson important to you?

- As a Datacap Solution/Application Builder, you build and deploy applications with the Datacap capture system and communicate solution details to the solution architect, administrator, and business analysts.
- To build an efficient Datacap application That saves users time, you need to be expert in the use of the validate actions because they are key to identifying data errors on captured documents.

Figure 2-66. Why is this lesson important to you?

IBM Training

Validate rule for the issue dates

- Add a rule to validate field data for both dates.
  - Is the field exactly 8 characters long?
  - Is the value a valid date in one of the standard date formats.

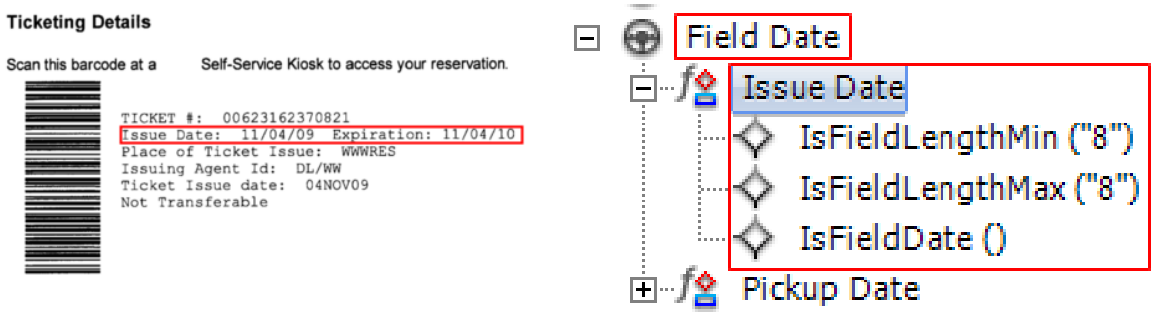


Figure 2-67. Validate rule for the issue dates

Images are:

- Page image format for the Issue and Expiration date
- Validate Date Rule

Validate date

How do you tell a valid date on the two Air_Receipt date fields?

If you are dealing with an eight-character format, 11/11/11, you can use the IsFieldLengthMax and IsFieldLengthMin actions. Then, you can use the IsFieldDate action to verify that the three components of the date are in range.

Another action that you can use to validate that a date is within a specific date range is the IsFieldDateWithinRange action. This check is not valid for the scenario that you are dealing with, so you are not going to use this action.

Depending on your requirements, you can control data tightly. For example, you can verify that the expiration date is not before the issue date.

For this demonstration, keep it simple, using only IsFieldLengthMax, IsFieldLengthMin, and IsFieldDate. The same rule is used for both date fields.

IBM Training

Validate rule for the ticket number field

- Add a rule to validate field data for the ticket number.
  - Does field contain any data?
  - Is the data limited to decimal digits?
  - Is the field size exactly 14 digits?

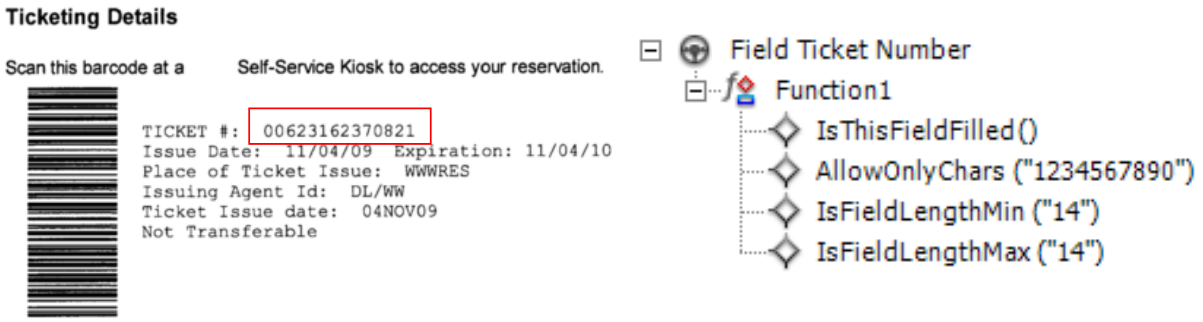


Figure 2-68. Validate rule for the ticket number field

Images are:

- Page image format for the Ticket number.
- Validate Ticket Number Rule

IBM Training

Verify

- Operator reviews pages to confirm and correct problem fields.
  - If validation errors were detected with Routing rules.
  - If verification is forced by using the Routing Force Verify rule.
- Documents with no errors can bypass operator review.
  - If, on the page, the DCO status is not set with the Routing Force verify rule.
- You can fill empty fields by the click and key method.
  - If a full page OCR operation is done on the page.
  - To do click and key, click the field on the image next to the form.
- Correct erroneous data.
  - Manually verify image data and retype it on the form.
- Manually verify all fields on the form.
  - Uses Alt V to force verification after changes.

Figure 2-69. Verify

Overview

During verification, Datacap shows pages to an operator to confirm, and if necessary, correct problem fields. You have control over whether documents with no validation errors are shown. Problem fields include:

- Character fields with one or more low-confidence characters
- OMR fields with low-confidence values
- Fields with validation errors

IBM Training

Exercise: Expand the validate rules

Basic Learning Application

Figure 2-70. Exercise: Expand the validate rules

IBM Training

Exercise objectives

- Expand the validate rules for the Air_Receipt pages

Figure 2-71. Exercise objectives

Lesson 2.7. Export to an IBM FileNet repository

IBM Training

Export to an IBM FileNet repository

Basic Learning Application

Figure 2-72. Export to an IBM FileNet repository

IBM Training

Lessons

- Create a Learning Template application
- Configure a Learning Template application
- Create locate rules
- Create validate rules
- Expand the locate rules
- Expand the validate rules
- Export to an IBM FileNet repository

Figure 2-73. Lessons

IBM Training

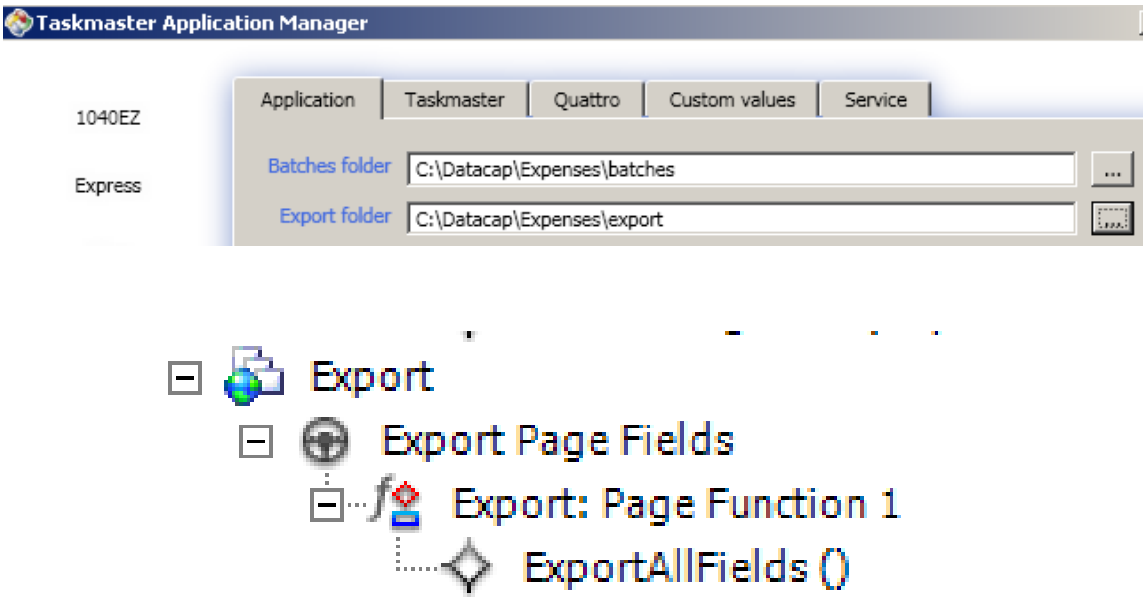
Why is this lesson important to you?

- As a Datacap Solution/Application Builder, you build and deploy applications with the Datacap capture system and communicate solution details to the solution architect, administrator, and business analysts.
- To build a complete Datacap application, you must be able to complete the cycle by exporting the final captured documents to the selected content store. In this lesson, you export to an IBM FileNet P8 repository.

Figure 2-74. Why is this lesson important to you?

IBM Training

Export types

- Export data to a text file, an XML file, a database, a document management system, or a custom business process.
- Default Export Folder - Exports folder for the application.
- 

Basic Learning Application

Figure 2-75. Export types

The images on the slide show:

- The Export folder parameter in the Datacap Application Manager.
- The Export ruleset that is used to export all images to text file.

Overview

The Export step is the last step of the capture process. This step determines where the final capture data is placed and what form it is in when it is exported.

Export to a text file

Datacap can export data to a text file, an XML file, a database, a document management system, or a custom business process. The default output format is a text file. The text file is the form that you are going to use in this section. Other export data forms are introduced in later sections of this class. The default location to place the output is in the exports folder for the application. Another location, the FileNet P8 repository, is introduced in a later lesson.

Default Export Folder

The default export folder is defined in the Datacap Application Manager by the Export folder parameter.

The default Export Page Fields rule stipulates that all identified fields are exported.

IBM Training

Preparation for export to IBM FileNet repository

- Create a folder structure to receive diverse document types if that is what you are exporting.
- Define a document class for each document type.
  - The document class requires a property that is defined for each field that is extracted from the captured document.
  - If the same field data types are extracted from all of the image types, then the same document class can be used to store the documents on the IBM FileNet repository.
    - The Expenses DocClass is for both the Rental_Agreement and Air_receipt in the lab.
    - You might add a property that indicates document type, that is, Rental_Agreement or Air_Receipt but the labs don't do that.
  - It is more likely that different document classes are preferred for each captured document type.

Figure 2-76. Preparation for export to IBM FileNet repository

Overview

Datacap applications can export captured data and images to a flat text file or to other external repositories with standard actions and separately licensed connectors. The external repositories include: systems and databases, IBM repositories, and other vendor repositories. In an earlier lesson, data was exported to a text file. In this exercise, you set up rules to export to an IBM FileNet repository.

Some preparation must be done for exporting to an IBM FileNet repository. Define a document class for a set of page data that is written to the object store. The document class must have a property that is defined for each value that is extracted from the page. Define three export rules. One rule is a batch rule that establishes a connection to the object store that the export data is written to. The other IBM FileNet export rules are one for each of the document sets that the Expense application processes.

Actions to export to IBM FileNet repository

FileNetP8 Action	Description
FNP8_Login	Sets the user ID and password for login to the FileNet P8 system.
FNP8_SetDestinationFolder	Sets the destination folder for the documents that are uploaded.
FNP8_SetTargetClassID	Sets the location type where docs will be stored, that is, ObjectStore
FNP8_SetTargetObjectID	Sets the FileNet P8 object store for the uploaded files.
FNP8_SetLocale	Sets the locale, that is, en_US
FNP8_SetURL	Sets the URL for the FileNet P8 web service.

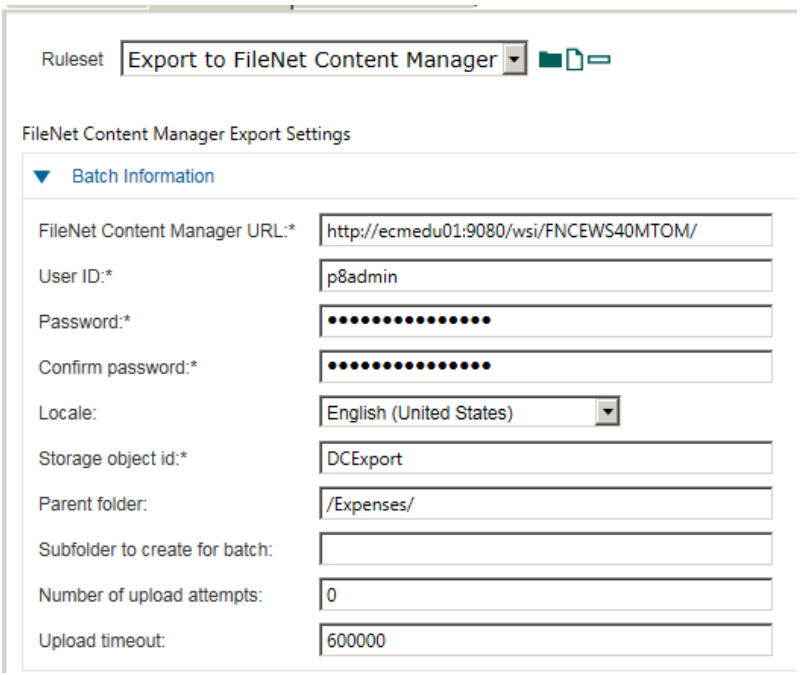
Figure 2-77. Actions to export to IBM FileNet repository

The actions for exporting data to a FileNet P8 repository are in the FileNet P8 group.

IBM Training

FastDoc batch level configuration

- Learning and Form Template both use the Export to FileNet Content Manager ruleset.



Basic Learning Application

Figure 2-78. FastDoc batch level configuration

The images show:

- The FastDoc interface for batch level configuration for the Export to FileNet Content Manager ruleset.
- The Datacap Studio definition representation of the document level configuration.

FileNet Content Management URL

- The URL for remote access to the IBM FileNet Content Manager Repository

User ID

- The user id used to log in the IBM FileNet Content Manager server.

Password and Confirm Password

- The password used to log in the IBM FileNet Content Manager server.

Storage object id

- The repository name

Parent folder

- The folder where Datacap documents are exported to.

IBM Training

Function to connect to the IBM FileNet server

- Learning and Form Template default batch level rule.
  - It shows configuration for the student image lab exercise.

Figure 2-79. Function to connect to the IBM FileNet server

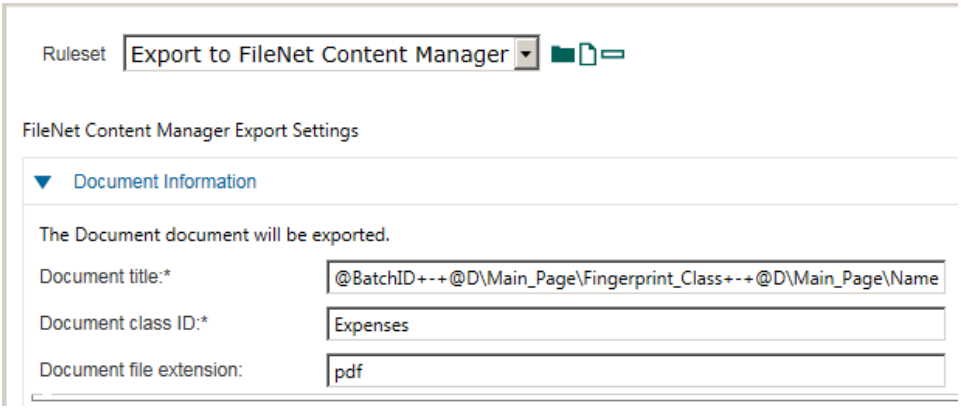
The images show:

- The Datacap Studio representation of the document level configuration.

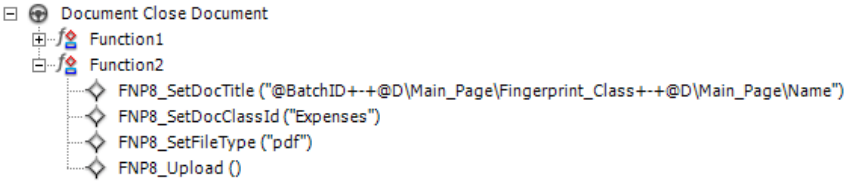
IBM Training

FastDoc document level configuration

- Learning and Form Template document level configuration the Export to FileNet Manager ruleset.



- Datacap Studio view of document level configuration



Basic Learning Application © Copyright IBM Corporation 2017

Figure 2-80. FastDoc document level configuration

The images show:

- The FastDoc interface for document level configuration.
- The Datacap Studio representation of the document level configuration.

Document title parameter analysis

- @BatchID - The same as the folder name in the batches folder. Example 20160829-000004.01
- @D\Main_Page\<field name> - this example provides access any field in the Main_Page page.
- Fingerprint_Class examples for this class are Rental_Agreement and AIR_Receipt.

Document Class ID

- Is the document class as defined in the IBM FileNet repository.

Document file extension

- The combination of the Batch ID and the Document file extension.
- The document must already be conversion to the file type that you specify in this parameter.
- In the lab exercise for this lesson, you use the CreateTIFF or PDF ruleset to produce a pdf to document to export.

IBM Training

FastDoc field level configuration

- For each document property to write to IBM FileNet repository backend server, you must configure a field level rule for the Export to FileNet Content Manager ruleset.

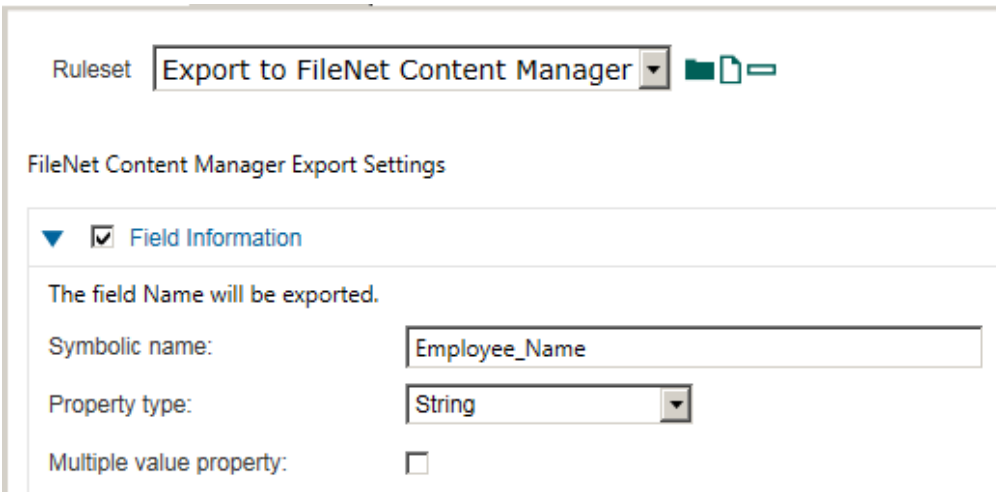


Figure 2-81. FastDoc field level configuration

The images show:

- The FastDoc interface for field level configuration.

Symbolic name:

- The name of the document property that is defined in the IBM FileNet repository Document Class.

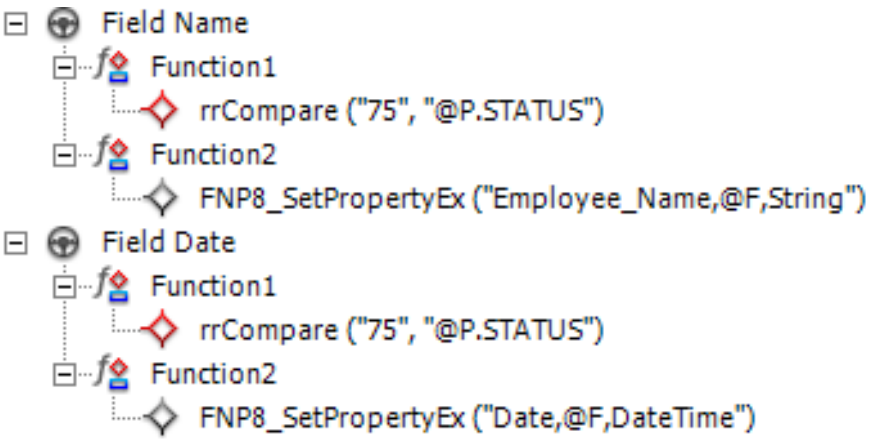
Property Type:

- This type that is defined for the Datacap field type must match the type of the property that is defined in the IBM FileNet repository Document class.

IBM Training

Setting the value of a document's properties

- Function to set field property values
- You must specify the field name on the IBM FileNet repository server and the data type
- In addition, you must pass in the field value



Basic Learning Application

Figure 2-82. Setting the value of a document's properties

The images show:

- The Datacap Studio representation of the field level configuration.

Definition for the **FNP8_SetPropertyEx** (StrParam) action.

Parameters: comma-separated string values:

1. Property ID: The name of an existing document property in the FileNet library (equivalent to a document index field)
2. String value or a predefined Smart Parameter variable to assign to the property. @F represents the field that this rule is associated with.
3. Optional property type: If this parameter is not specified, the property type defaults to a 'string'. Supported types are Binary, Boolean, DateTime, Float, ID, Integer, Object, and String.

IBM Training

Configure multi-page documents

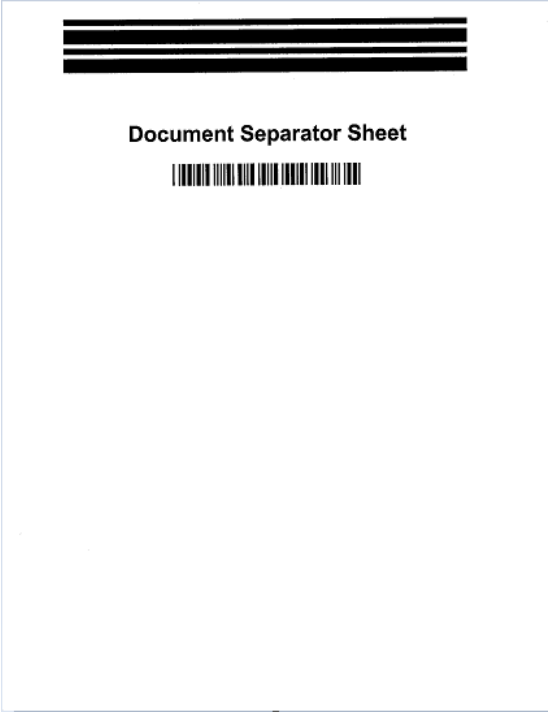
- Learning Template assumes one page per document.
- To group multiple pages into a document use separators page.
- General rule:
  - Pages are treated as one page documents until a separator page it encountered.
  - Then all pages are assumed to be in the same document until another separator page is encountered.
- Sample multi page batch:
  - Separator page.tif
  - Car1.tif
  - Air1 sample1.tif
  - Hotel1.tif
  - Separator page.tif

Figure 2-83. Configure multi-page documents

IBM Training

Separator page sample

- Adjust Image Enhancement
- Set the Remove Lines > Minimum



Basic Learning Application

Figure 2-84. Separator page sample

The image shows a sample separator page with a horizontal bar, The words ”Document Separator Sheet”, and a vertical bar code.

- When you first configure your multi-page batch for a Learning Template application, you might detect the following behavior:
  - The separator pages are treated as standard page.
- The cause of this is probably because the Separator page bare code is being corrupted by the Image Enhancement rulset.
- To correct this condition, set the Minimum length parameter under the Remove Lines option to a value >= 70.
- Expense > Document > Main_Page > Image Enhancement > Remove Lines > Minimum length

IBM Training

Demonstrations

- Export to IBM FileNet Content Manager

Figure 2-85. Demonstrations

If you are taking this course as a self-paced virtual class, return to the main course menu to play the pre-recorded demonstrations.

IBM Training

Exercise: Export to an IBM FileNet repository

Basic Learning Application

Figure 2-86. Exercise: Export to an IBM FileNet repository

IBM Training

Exercise objectives

- Export Expense documents to an IBM FileNet repository

Figure 2-87. Exercise objectives

IBM Training

Unit summary

- After completing this unit, you should be able to:
- Create a basic learning application with the Datacap Application wizard
- Apply image enhancement techniques and field recognition potential
- Define a document structure
- Create and expand Locate rules
- Create and expand Validate rules
- Export the scanned documents to an IBM FileNet repository

Figure 2-88. Unit summary

Unit 3. Debug and Test

Estimated time

02:00

Overview

This unit introduces you to the tools and interfaces that are used for debugging and testing Datacap applications.

How you will check your progress

- Successfully complete the activities in the Student Exercises book.

References

IBM Knowledge Center

http://www.ibm.com/support/knowledgecenter/SSZRWV_9.0.1/com.ibm.datacaptoc.doc/datacap_9.0.1.htm

IBM Training

Unit Objectives

- After completing this unit, you should be able to:
- Configure and access logs for debugging
- Use the Datacap Studio Test tab to step through a batch for testing

Figure 3-1. Unit Objectives

Lesson 3.1. Application Debugging

Figure 3-2. Application Debugging

IBM Training

Topics

- Application Debugging
- Datacap Studio Test Tab

Figure 3-3. Topics

IBM Training

Why is this lesson important to you?

- As a Datacap business analyst, you build and deploy applications with the Datacap Capture system and communicate solution details to the solution architect, administrator, and users. As part of that process, you must be able to debug solutions.
- You can run an application from Datacap Studio and monitor it during execution, then determine whether the rules are running as you expect.
  - Become familiar with the debugging features that are included with the product by intentionally creating errors in your application.

Figure 3-4. Why is this lesson important to you?

IBM Training

Application Debugging and Datacap Studio

- Datacap Studio includes integrated debugging functionality.
  - Through this functionality you can control execution of your application.
- You can run an application from Datacap Studio.
  - You're able to monitor the application during execution.
  - You can also determine whether the rules are running as you expect.
- Application debugging requires that you review at least two runtime log files.
  - The Rulerunner Service (RRS) log.
  - The Task Log.
- There are also a few other logs that are useful depending on what the issue is with your application.

Figure 3-5. Application Debugging and Datacap Studio

Application debugging requires that you review two runtime log files, which are the Rulerunner Service (RRS) log and the task log. The RRS log provides detailed information about each action and is most helpful to application developers. The task log documents internal calls and is used mostly by IBM software support.

Datacap Studio includes integrated debugging functionality through which you can control the execution environment and monitor your application at runtime.

IBM Training

The Logs Used in Debugging

- As Rulerunner runs each action, it writes detailed logging information to a Rulerunner Service (RRS) log file.
- The Rulerunner Service Logs = RRS Logs:
  - They provide detailed information about each action as it is executed.
  - These logs are most helpful to application Developers.
  - There is a global RRS log for each execution of an application as well as individual RRS logs for components like TM Server, Rulerunner server, and TM Web.
- The Task Log:
  - Documents internal calls.
  - Is used mostly by IBM software support.

Debug and Test

Figure 3-6. The Logs Used in Debugging

Datacap log files

Datacap generates two types of log files during task execution.

- Rulerunner Service (RRS) log files include detailed information about each action as it runs.
- Task log files document mostly internal calls and is most helpful to IBM Software support.

Additionally, Report Viewer and Rulerunner can generate their own log files.

IBM Training

The RRS Log Files

- Rulerunner also generates an RRS log file whenever you run a task from Datacap Studio.
  - RRS task log files are named task_rrs.log, ie. verify_rrs.log.
- You can set Rulerunner logging by application and task.
- Logging is enabled by default for all tasks except Vscan and tasks that you start from Datacap Studio.
  - Each log file contains detailed descriptions of the actions that are run by the task profile.
  - Logging is useful for application troubleshooting.

Figure 3-7. The RRS Log Files

Rulerunner Service (RRS) log files

As Rulerunner runs each action, it writes detailed logging information to a Rulerunner Service (RRS) log file (*task*_rrs.log). Rulerunner also generates an RRS log file whenever you run a task from Datacap Studio.

1. If you want to generate an RRS log file for tasks that you run from the Datacap Web Client or for Datacap Desktop tasks, complete the following steps.Start Datacap Rulerunner Manager.
2. Click the Logging tab.
3. Click the RRS log tab and select the logging options that you want.

In the Datacap Web Client, each task generates its own Rulerunner Service log file. The most recent TravelDocs batches folder contains a log file for each of the task profiles in the Main Job workflow.

Each log file contains detailed descriptions of the actions that are run by the task profile and is useful for application troubleshooting.

IBM Training

Web Client and Datacap Desktop Logging

- Logging of tasks run from either the Datacap Web Client, or Datacap Desktop can be configured
- To log from either or both clients, complete the following steps:
  - Start Datacap Rulerunner Manager.
  - Click the Logging tab.
  - Click the RRS log tab and select the logging options that you want.
- In the web client, each task generates its own Rulerunner Service log file.
- Each log file contains detailed descriptions of the actions run by the task profile.
  - Useful for application troubleshooting.
  - There are good examples in Datacap 9.0 help.

Debug and Test

Figure 3-8. Web Client and Datacap Desktop Logging

Enable logging for Datacap Web Client tasks

To enable logging for a web task, you must configure that task in the Datacap Web Client.

1. To enable logging from web client tasks:In the Datacap Web Client, click the Administrator tab.
2. On the Administrator tab, click Workflow.
3. Expand the job that contains the task for which you want to enable logging, and select the task.
4. Click Setup in the Selected task details pane.
5. In the Rulerunner service log field, enter one of the values, as required.Tip: RRS logging is only useful for tasks that run rules. If your web client is not associated with a task profile, an RRS log file is not generated.

In most situations, a setting of 3 provides enough information to help you debug rule-related issues.

IBM Training

Setting Rulerunner Logging by Application/Task

- To configure Rulerunner logging by application and task, string values must be added to the Registry.
- Steps to enable RRS logging by application name:
  - Add a String Value with the Value name app_filter to the registry.
  - Then enter the task name in the Value Data field.
  - For example, if you want to enable logging for the APT application tasks, add the app_filter value name and enter APT in the Value Data field.
- You can also enable RRS logging by task profile.
- Detailed information is in the Datacap 9.0 online help.

Debug and Test

Figure 3-9. Setting Rulerunner Logging by Application/Task

Set Rulerunner logging by application and task

You can set Rulerunner logging by application and by task when you add string values to the registry.

To enable RRS logging by application name, add a String Value with the Value name app_filter to the registry. Then, enter the application name in the Value data field. For example, if you want to enable logging for the APT application tasks, add the app_filter value name and enter APT in the Value data field.

- To enable RRS logging by task profile, add a String Value with the Value name tprofile_filter to the registry. Then, enter the task name in the Value data field. For example, if you want to enable logging for the VScan task, add the tprofile_filter value name and enter VScan in the Value data field.
  - For 32-bit (x86) OS, add the value in the following location.
  - HKEY_LOCAL_MACHINE\SOFTWARE\Datcap\Rulerunner\RRSLog
  - For 64-bit (x64) OS, add the value in the following location.
  - HKEY_LOCAL_MACHINE\SOFTWARE\Wow6432Node\Datcap\Rulerunner\RRSLog

You can set logging for an application and a task profile by adding the app_filter and the tprofile_filter string values to the registry. For example, if you set the app_filter value to APT and the tprofile_filter value to VScan, RRS logging is enabled only when the VScan task profile in the APT application is run.

If the app_filter Value data field is empty, RRS logging is enabled for all applications. If the tprofile_filter Value data field is empty, RRS logging is enabled for all task profiles.

- To confirm that the registry log filter is enabled, look in the Rulerunner.log file or look in the batch. The Rulerunner.log file indicates if application or task filtering is enabled for RRS logging.
- The batch has only the .rrs log files for the filtered task. For example, if you set the tprofile_filter value to VScan, then only the VScan.rrs logs are in the batch.

Note: You must restart the Rulerunner Service when you add a registry key log filter or change the value of a log filter.

IBM Training

Datacap Web Client Tasks and Log Files

- Logging is enabled by default for all tasks except Vscan and tasks that you start from Datacap Studio.
- If you want to generate a task log for Datacap Web Client tasks, you must set the severity level to 1 or higher (on a scale of 0-8).
  - The severity level and the options in the Datacap Web Client Task setup window determine how much information is written to the log file.
- The Task Log file is saved in the batch folder.
  - It provides information that is typically most helpful to IBM support because it documents mostly internal calls.

Figure 3-10. Datacap Web Client Tasks and Log Files

IBM Training

Support Recommendations

- One of the best ways to debug an application is to read the relevant log files after asking yourself “What is failing?”
  - Become familiar with the information that is in the various log files so that when you run into a problem, you’re already familiar with the contents of the files and the kinds of information they contain for non-problematic batches.
  - When a problem occurs, ask things like whether if fails running manually? Is it working outside of Rulerunner server? Are you using a client application (ie. DotScan)? Are there issues with TM server? If so, gather the tmserver log.
- Be aware of what is logged in the Rulerunner log file.
  - Also be aware of the information in the rrs log for ie. a correctly run Page ID step in your workflow.
  - That will be helpful so that you recognize lines of information that might be problem related when Page ID has an error.

Figure 3-11. Support Recommendations

IBM Training

Further Recommendations

- Before you run into problems, make sure you know how to configure and use the debugging tools.
- Search in the extensive help which ships with the product.
- A couple of extremely helpful searches related to logs are the following:
  - http://www.ibm.com/Search/?q=datacap+logging&v=17&en=utf&lang=en&cc=us
  - http://www-01.ibm.com/support/knowledgecenter/SSZRWV_8.1.0/com.ibm.dc.admin.doc/dclog005.htm
- Notice that there is a reference to the Datacap 8.1 knowledge center.
  - Much of Datacap 8.1 help is relevant to Datacap 9.0.
  - There is much more 8.1 debug information because Datacap 9.0 is so new.

Figure 3-12. Further Recommendations

IBM Training

Another Key Way to Debug

- Another excellent way to debug applications is to use the Test Tab in Datacap Studio – covered in last lesson.
- An incomplete list of things you can do using the Test Tab:
  - Set breakpoints.
    - Breakpoints stop a task at a predetermined rule set, rule, or action.
    - You can also configure a breakpoint that stops a task when it starts processing a specific document, page, or field.
  - Single step through your code.
    - Useful to determine whether the functions and actions within a rule are operating as intended.
    - As you step through each line, you can see the actions that are returned as True (check mark) and False (exclamation point).
  - View output that is written to different log files via the Output tab in Datacap Studio.

Figure 3-13. Another Key Way to Debug

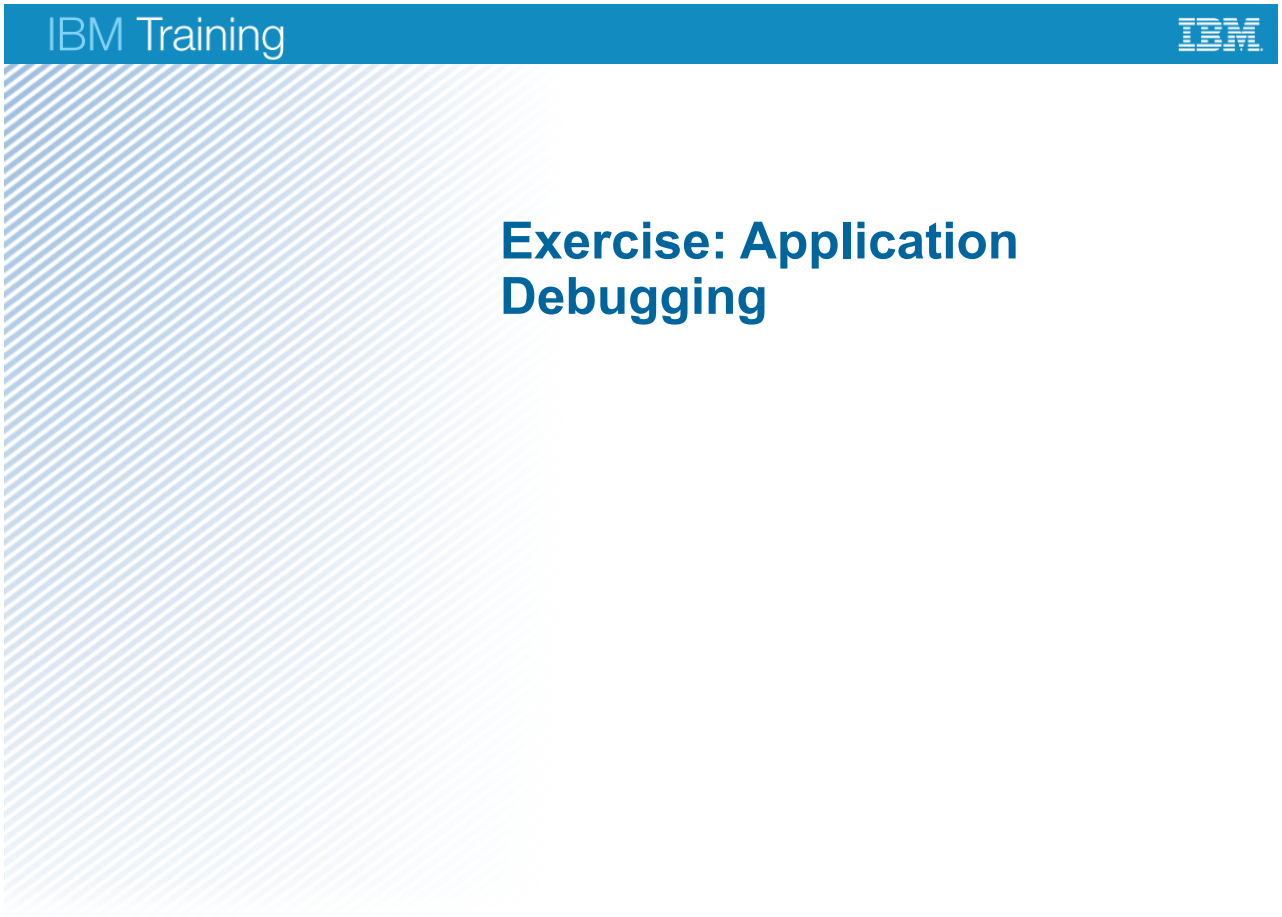


Figure 3-14. Exercise: Application Debugging

IBM Training

Exercise objectives

- Configure and access logs for debugging

Figure 3-15. Exercise objectives

Lesson 3.2. Datacap Studio Test Tab

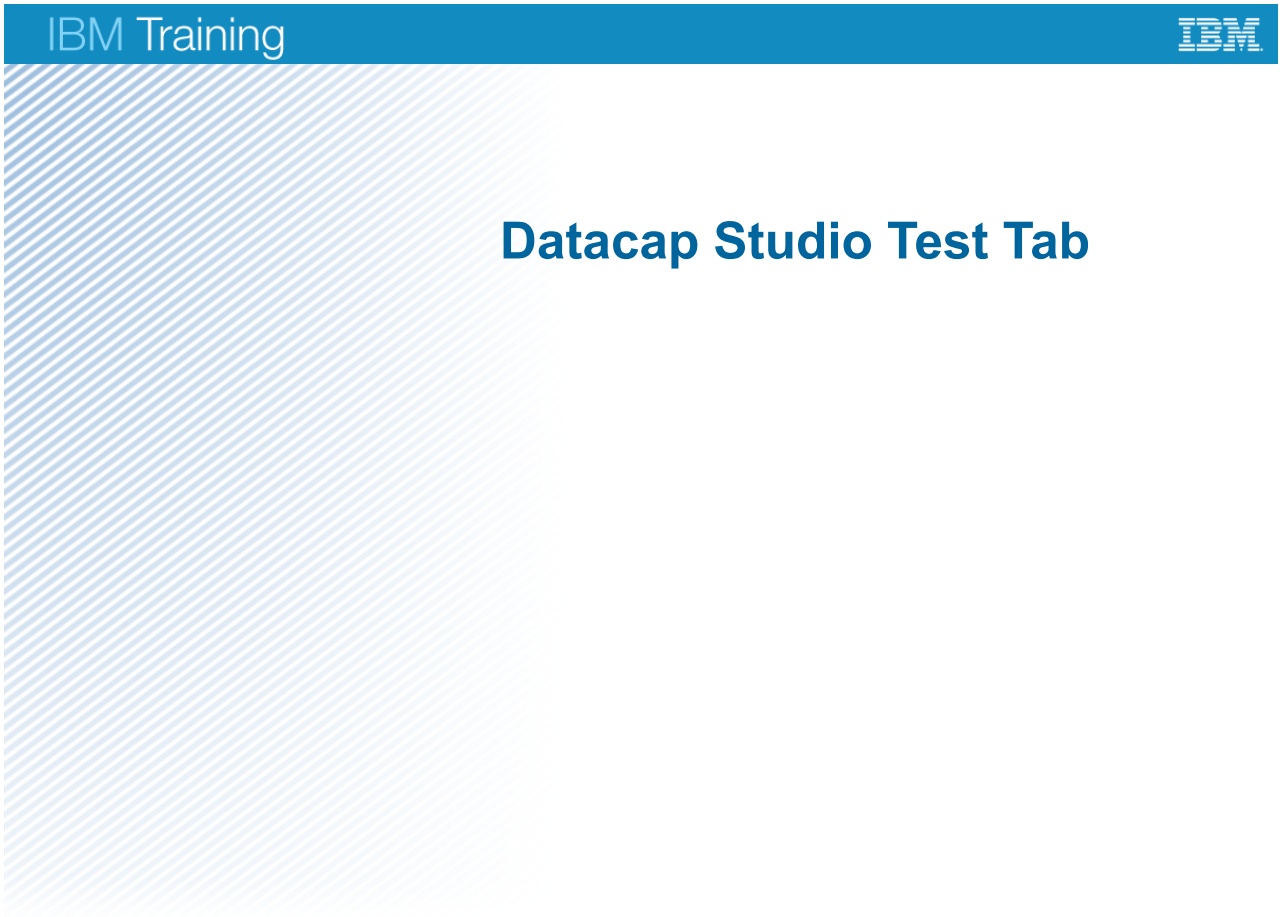


Figure 3-16. Datacap Studio Test Tab

IBM Training

Topics

- Application Debugging
- Datacap Studio Test Tab

Figure 3-17. Topics

IBM Training

Why is this lesson important to you?

- As a Datacap business analyst, you build and deploy applications with the Datacap Capture system and communicate solution details to the solution architect, administrator, and users. As part of that, you must be able to debug solutions.
  - To configure and test most Datacap capture applications, it's very helpful, and speeds the development cycle, if the developer makes use of the Datacap Studio Test Tab.
    - The Datacap Studio Test Tab includes debugging features.
  - You can run an application from Datacap Studio to monitor it during execution and determine whether the rules are running as you expect.

Figure 3-18. Why is this lesson important to you?

IBM Training

Datacap Studio Test Tab and Debugging

- The Datacap Studio test tab contains 8 panels:
  - Workflow
  - Runtime batch hierarchy
  - Document hierarchy
  - Rulesets
  - Image/Text
  - Batch Data
  - Properties
  - Breakpoints/Runtime state/Call stack

Figure 3-19. Datacap Studio Test Tab and Debugging

Datacap Studio Test Tab

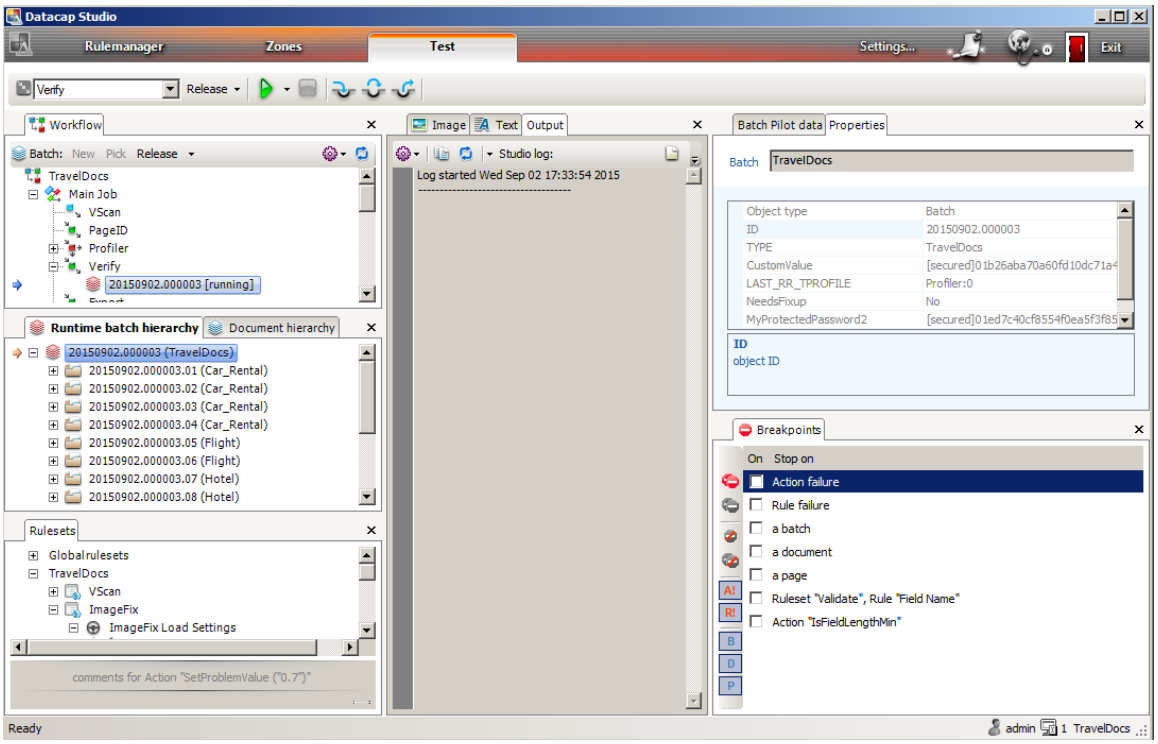


Figure 3-20. Datacap Studio Test Tab

The image on the slide illustrates the Test view of Datacap Studio

- Control Bar
- Workflow pane
- Hierarchy Pane
- Ruleset pane
- Image pane
- BatchPilot data
- Properties
- Breakpoints

IBM Training

Four of the Eight Panels

- Workflow
  - Displays the job types and tasks defined in the Administrator tab.
  - Can run a batch through the workflow.
- Runtime Batch Hierarchy
  - When a batch is running, displays runtime batch hierarchy, including any data values.
  - If you select a page object, the page is displayed in the image panel.
- Document Hierarchy
  - Displays the structure of the documents you are processing.
  - Shows how each element within the structure is processed.
- Rulesets
  - Displays the rules, functions, and actions that make up each ruleset.
  - As you step through the workflow, you can see the current execution point.

Figure 3-21. Four of the Eight Panels

IBM Training

Other Four Panels

- Image/Text
  - Displays the selected page in the runtime batch hierarchy.
- Batch Data
  - Displays batch level information for the batch that is running.
- Properties
  - Displays the properties for the selected document hierarchy or ruleset object (read only).
- Breakpoints / Runtime State / Call Stack
  - A breakpoint stops processing at a predetermined ruleset, rule, or action.
  - Or it stops the task when a task starts processing a specific document, page, or field.

Figure 3-22. Other Four Panels

IBM Training

Test Tab Debugging Features

- Using Breakpoints
  - A breakpoint stops at a predetermined rule set, rule, or action.
  - It might instead stop a task when it starts processing a specific document, page, or field.
- Single-stepping through your code
  - Single-stepping is useful to determine whether the functions and actions within a rule are operating as intended.
  - As you step through each line, you can see the actions that returned True (check mark) and False (exclamation point).
- Examining log files from the Test tab
  - You can access the Output tab in Datacap Studio to view output written to different log files.

Figure 3-23. Test Tab Debugging Features

IBM Training

Setting and Using Breakpoints

- There are two types of breakpoints, both of which halt execution when Rulerunner encounters the specified element.
  - A breakpoint halts execution when the Rulerunner execution manager encounters the specified element, regardless of context.
  - A full breakpoint halts execution when the Rulerunner execution manager encounters the specified element within the same context.

Figure 3-24. Setting and Using Breakpoints

IBM Training

Setting Generic Breakpoints

- Generic Breakpoints: The Datacap Test tab includes two controls that you can select to halt processing when any rule or action fails:

Control	Command	Description
A!	**Stop on a failed action**	**Click this button to add a generic breakpoint that halts processing whenever an action fails**
R!	**Stop on a failed rule**	**Click this button to add a generic breakpoint that halts processing whenever a rule fails**

Figure 3-25. Setting Generic Breakpoints

The image on the slide illustrates the generic breakpoint run options:

A! – Stop on a failed action

R! – Stop on a failed rule

IBM Training

Breakpoint Configuration Interfaces

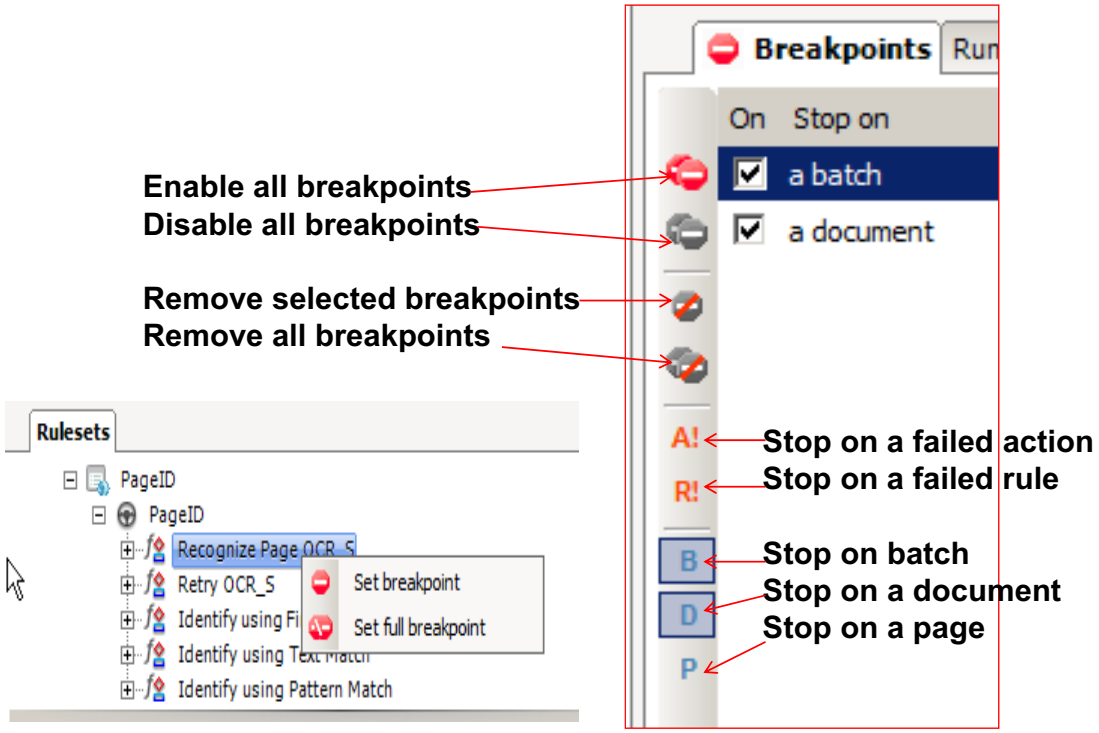


Figure 3-26. Breakpoint Configuration Interfaces

The image on the slide illustrates:

- Ruleset pane of the Test tab page
- Breakpoints pane of the Test tab page

IBM Training

Setting Breakpoints

- You can set a breakpoint for a rule set, rule, function, action, document, page, or field.
  - The document, page, or field must exist in the runtime hierarchy before you can set a breakpoint on it.
- To Set Breakpoints:
  - Set a breakpoint on a rule set, rule, function, or action:
    - Go to the rulesets pane on the Datacap Studio Test Tab.
    - Right-click the item and select Set breakpoint or Set full breakpoint.
  - Set a breakpoint on a document, page, or field:
    - Go to the Runtime batch hierarchy pane on the Datacap Studio Test tab.
    - Right-click the item and select Set breakpoint or Set full breakpoint.

Figure 3-27. Setting Breakpoints

IBM Training

Disabling and Clearing Breakpoints

- The Breakpoints pane displays all of the defined breakpoints.
- You can enable or disable individual breakpoints by selecting or clearing checkboxes.
  - The buttons on the left of the Breakpoints pane are options to enable, disable, or remove breakpoints.
- The check box to the left of each breakpoint indicates whether the breakpoint is enabled or disabled.
- By default, breakpoints are enabled when you add them.

Figure 3-28. Disabling and Clearing Breakpoints

IBM Training

Single Stepping Through Your Code

- Single stepping is useful to determine whether functions and actions within a rule are operating as intended.
- As you step through each line, you can see the actions that are returned as True (check mark) or False (exclamation point).
  - If an action returns false, you can look in the batch log to see why the action returned false.
- You can also access the Output tab in Datacap Studio to view the output that is written to different log files.

Figure 3-29. Single Stepping Through Your Code

IBM Training

How to Step Through Code: Step in

- The Test tab provides UI controls (enabled with tooltips) for stepping through code.
- Step in – steps into the next line of code. 
  - If the next line calls a rule or function, Step in opens the rule or function and halts inside it.
  - If the next line is an action, Step in opens the action; you must click it again to close the action.

Figure 3-30. How to Step Through Code: Step in

IBM Training

How to Step Through Code: Step Over or Out

- Step/Step over – starts the next line of code and any lower-level functions and actions, and then stops:
  - If the next line is an action, Step over works like Step in and opens the action.
- Step out – steps through the next line of code:
  - If the next line is a rule or function, Step out works like Step over and starts any lower-level functions and actions.
  - If the next line is an action, Step out starts and closes the action.

Figure 3-31. How to Step Through Code: Step Over or Out

IBM Training

Examining Log Files from the Test Tab

- You can access the Output tab in Datacap Studio to view the information that is written to different log files.
- 1. Click the Output tab in the center pane of the Test tab.
- 2. If Batch Log is not already selected, click the down-arrow and select it from the list of available logs.
- The output pane refreshes automatically when:
  - You stop at a breakpoint
  - You single-step through a line of code
  - When the current task profile completes
- You need to scroll to the bottom each time to see the latest messages.

Figure 3-32. Examining Log Files from the Test Tab

IBM Training

Demonstrations

- The Test tab in Datacap Studio

Figure 3-33. Demonstrations

If you are taking this course as a self-paced virtual class, return to the main course menu to play the pre-recorded demonstrations.

IBM Training

Exercise: Datacap Studio Test tab

Debug and Test

Figure 3-34. Exercise: Datacap Studio Test tab

IBM Training

Exercise objectives

- Stepping a batch through a PageID task profile

Figure 3-35. Exercise objectives

IBM Training

Unit summary

- After completing this unit, you should be able to:
- Configure and access logs for debugging
- Use the Datacap Studio Test tab to step through a batch for testing

Figure 3-36. Unit summary

IBM Training