

# Organize

## Overview

Thinking back to our real-world scenario, what might we need to do next if we were processing these documents by hand?

We probably don't want to throw all of our invoices into one giant file and call it a day, right? (Well, we might *want* to, but we have to at least pretend like we're responsible!)

If we were ultimately going to store these documents in a file cabinet somewhere, we would need to separate them using file folders so that one invoice doesn't get mixed up with another. We would also need to have some kind of label on the folders so we can easily identify what the folder contains and what information it has.

We have to do the same things with Grooper, and these are done in the `Organize` phase.

In this phase, we will:

- create a `Separation Profile`,
- use the new profile to separate our invoices into individual folders,
- learn about classification and content types,
- train Grooper to recognize our documents so it knows what information to get from them, and
- update our Batch Process to do all of this automatically in the future.

Let's get organized!

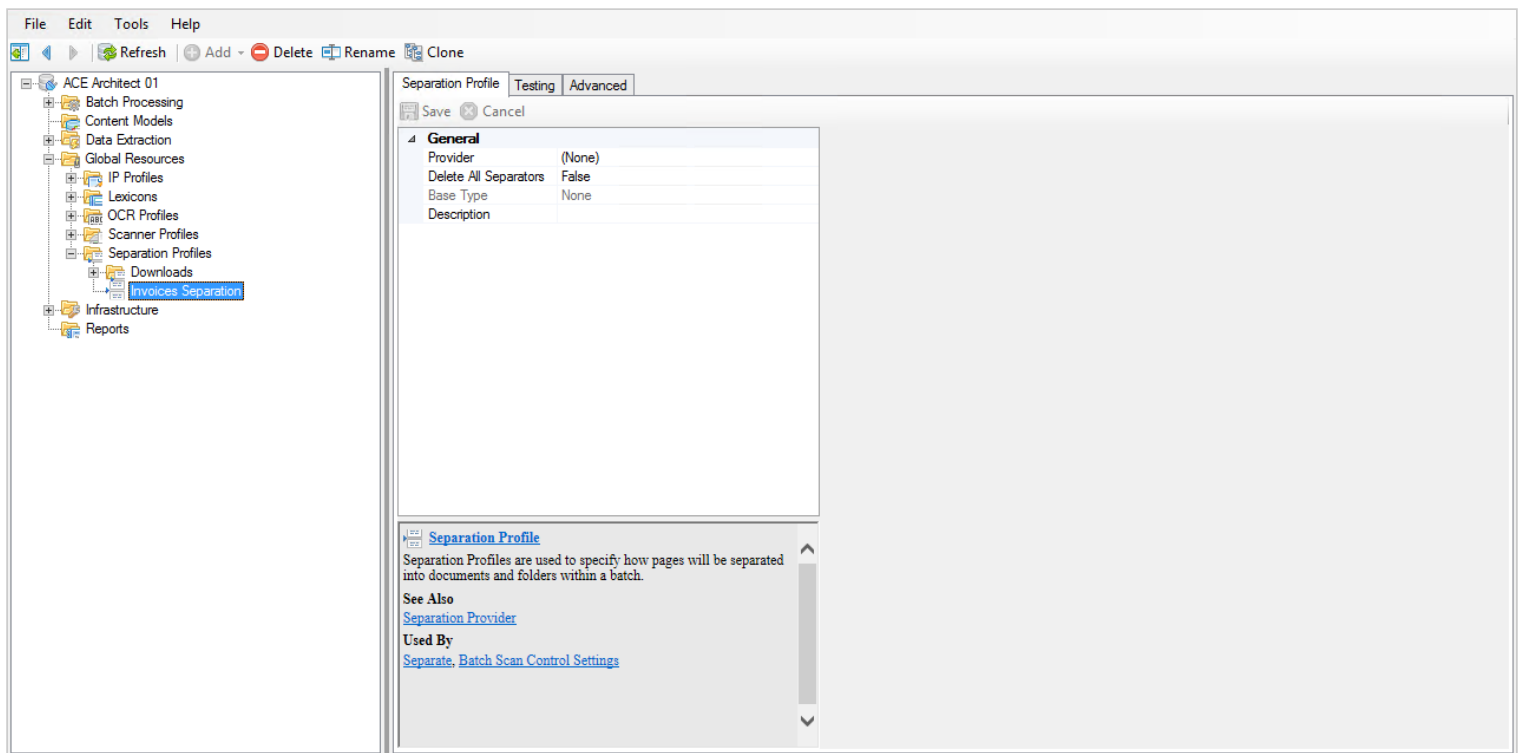
## Separation

Before we can expect to get any kind of information from these invoices, we need to sort them. While we could get away with doing this manually, Grooper most certainly provides us with a better way.

# Creating a Separation Profile

## > Step 1

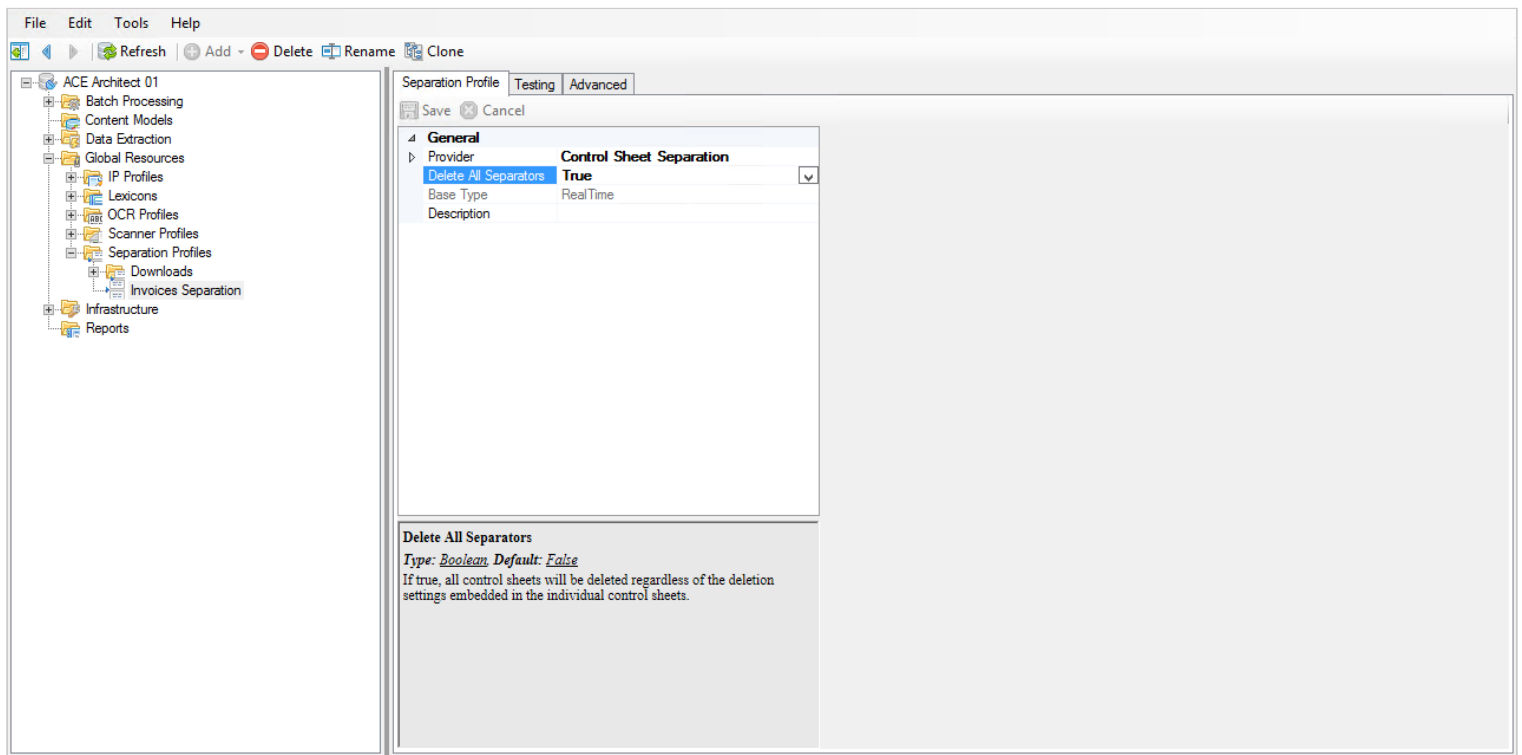
1. Navigate to (root) > Global Resources > Separation Profiles
2. Create a new separation profile
3. Give it a name, like Invoices Separation



## > Step 2

Change the following properties:

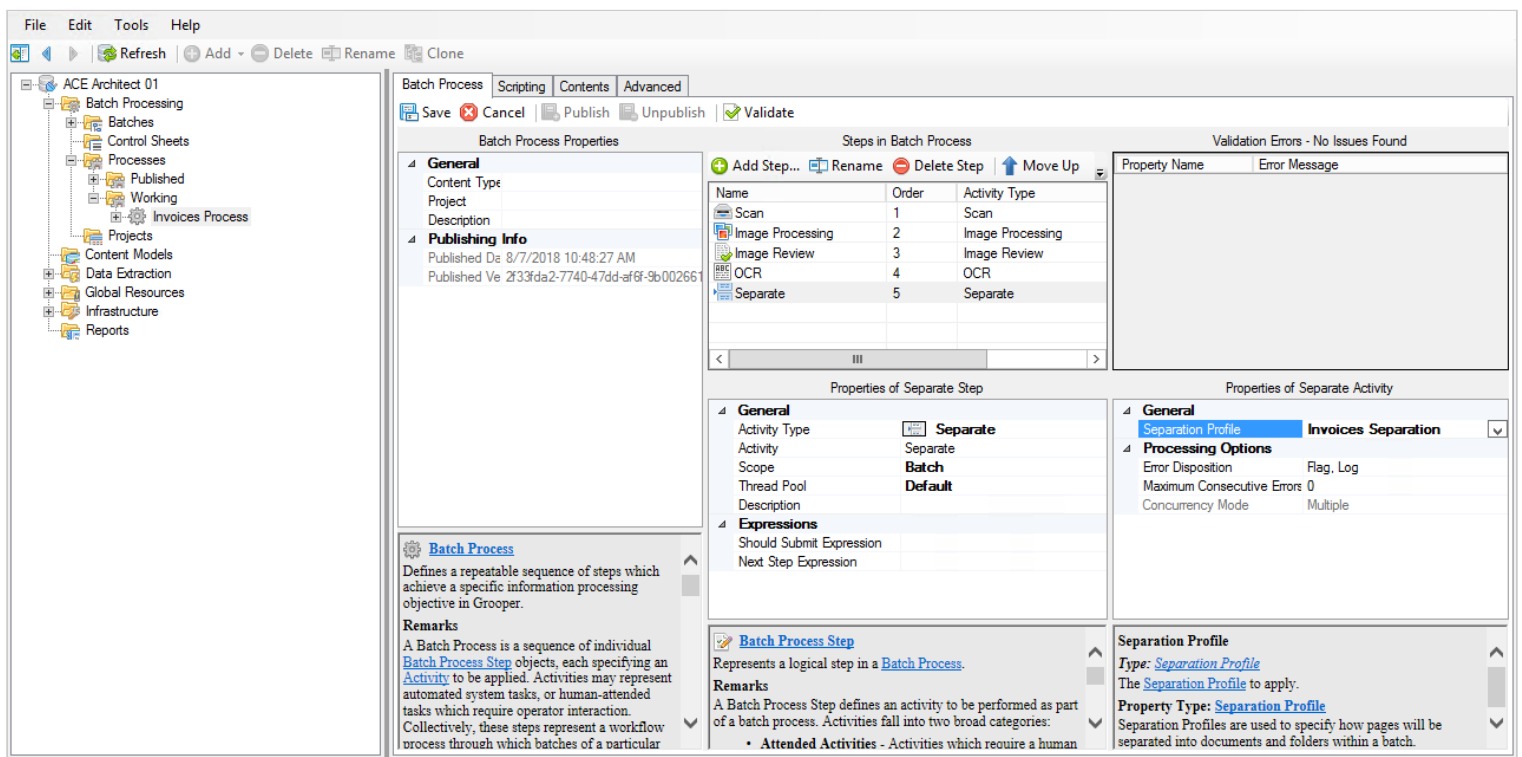
1. Set Provider to Control Sheet Separation
2. Set Delete All Separators to True



## Updating the production batch

### > Step 1

1. Navigate to (root) > Batch Processing > Processes > Working > Invoices Process
2. Add a new step
3. Set the Activity Type property of the new step to Separate
4. Set the Separation Profile to the new Invoices Separation profile we created
5. Save and Publish.



## > Step 2

1. Navigate to (root) > Batch Processing > Batches > Production > Invoices Process.
2. Make sure the batch is paused
3. Right click on the Invoices Batch and select Update Process....
4. Set Target Step to Separate.
5. Click Execute.

File Edit Tools Help

Refresh Add Delete Rename Clone

Folder Contents Advanced

1 / 1 batches selected Search:

Batch Name	Priority	Status	Batch Process	Current Step	Created
Invoices Batch			Invoices Process	(none)	8/6/2018 4:59:16 PM

Clone To Test  
Reset...  
Update Process...  
Share  
Refresh F5  
Select All Ctrl+A  
Delete Del  
Properties...  
New... Ins

Updates or changes the Batch Process associated with this batch.

Processing Status Statistics Tasks Batch Viewer Processing History Event Log Details

### Task Status

Legend: Pending (grey), In Progress (green), Completed (blue), Errors (red)

## Update Process

Execute Cancel

**General**

Target Step **Separate**

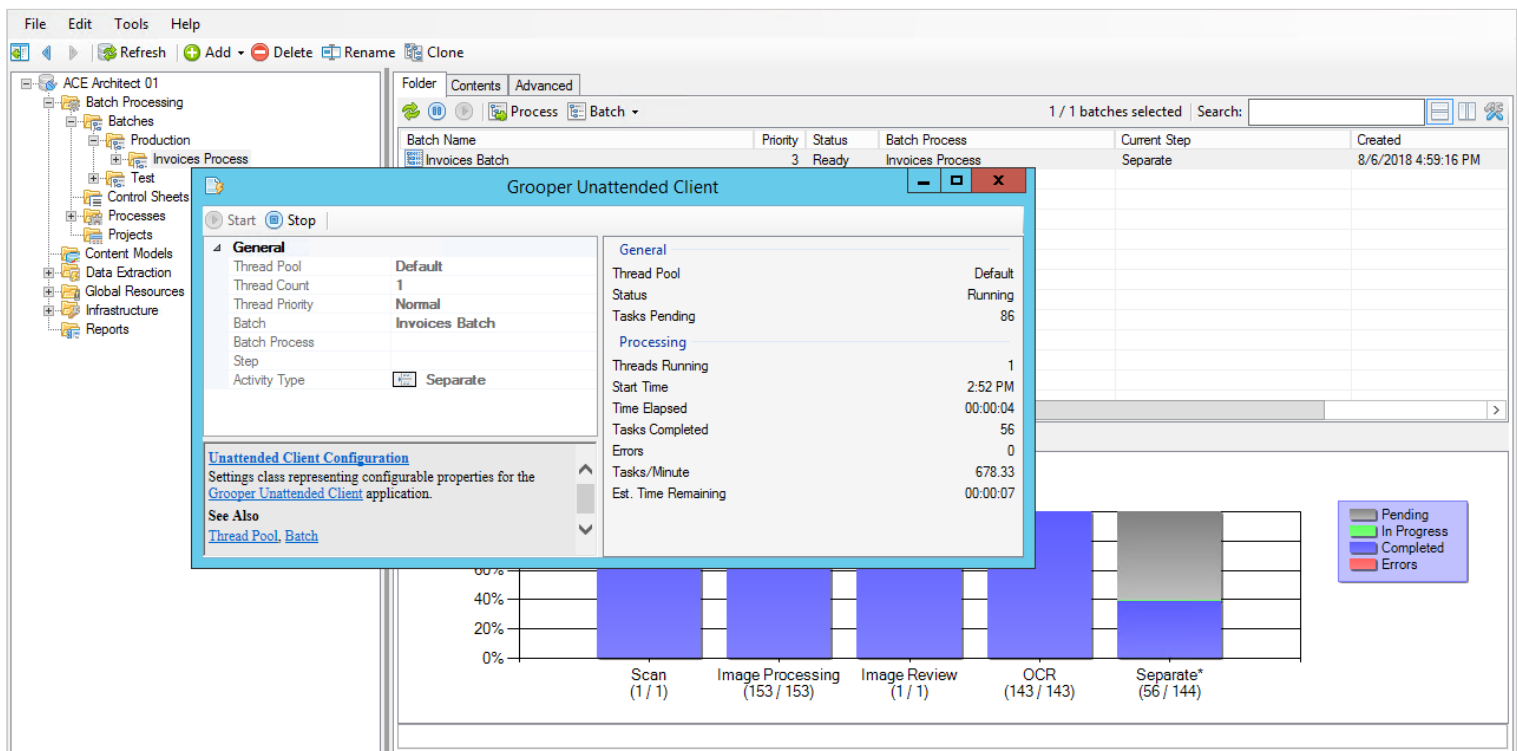
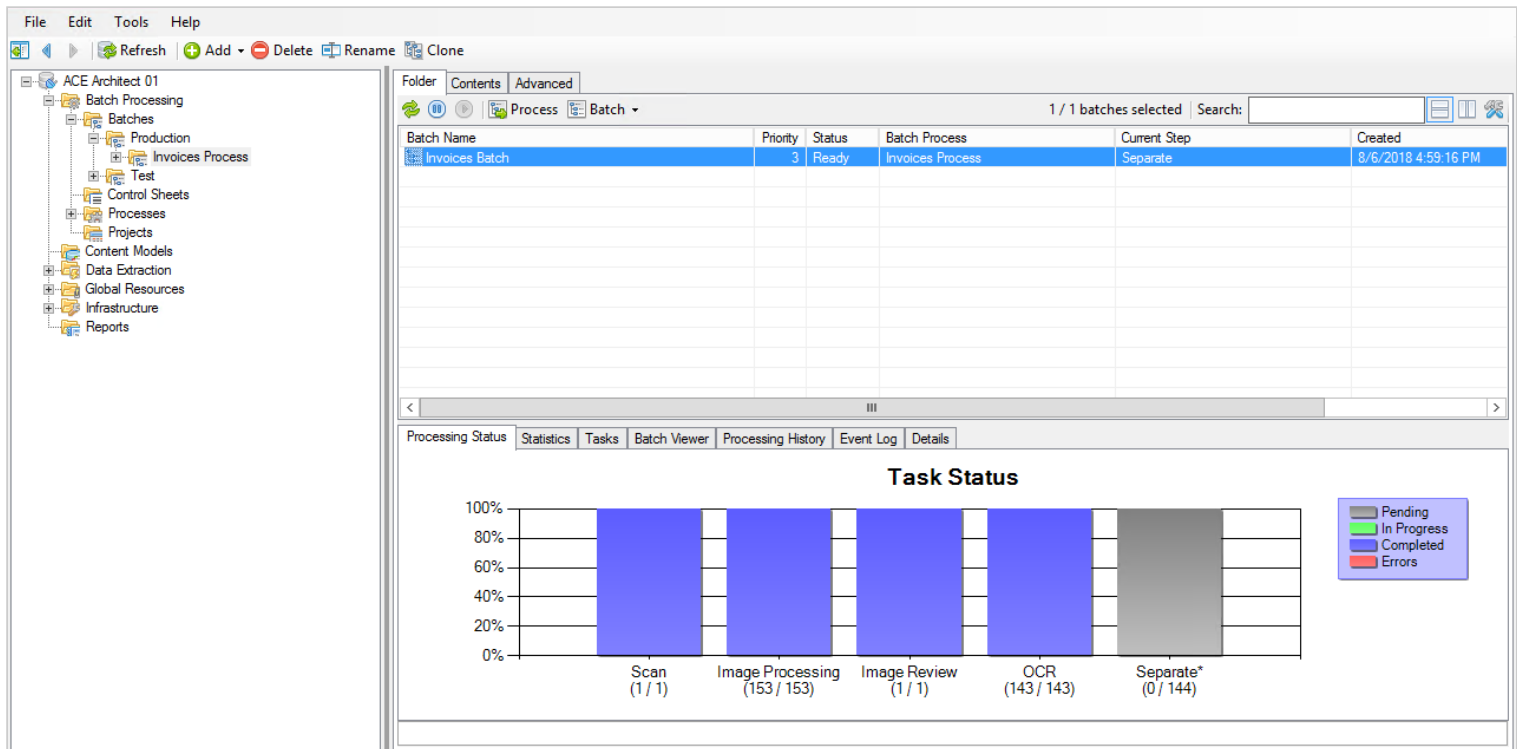
**Target Step**  
 Type: [Batch Process Step](#)  
 The [Batch Process Step](#) where the batch should begin processing.

**Remarks**  
 The batch will be updated to use the latest published version of the parent [Batch Process](#).

**Property Type: [Batch Process Step](#)**  
 Represents a logical step in a [Batch Process](#). A Batch Process Step defines an activity to be performed as part of a batch process. Activities fall into three broad categories:

### > Step 3

1. Resume the batch
2. Click Process

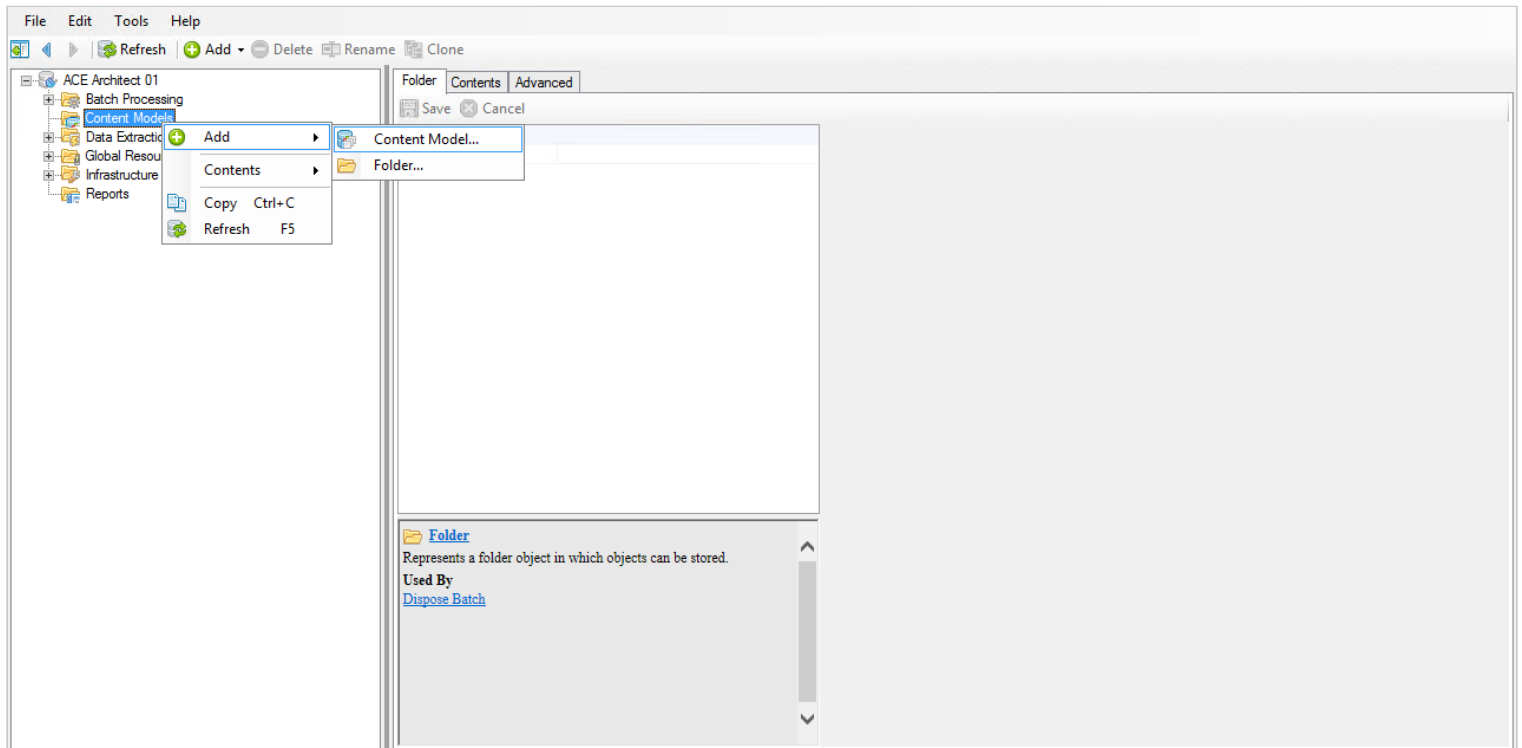


# Classification

## Setting up the Content Model

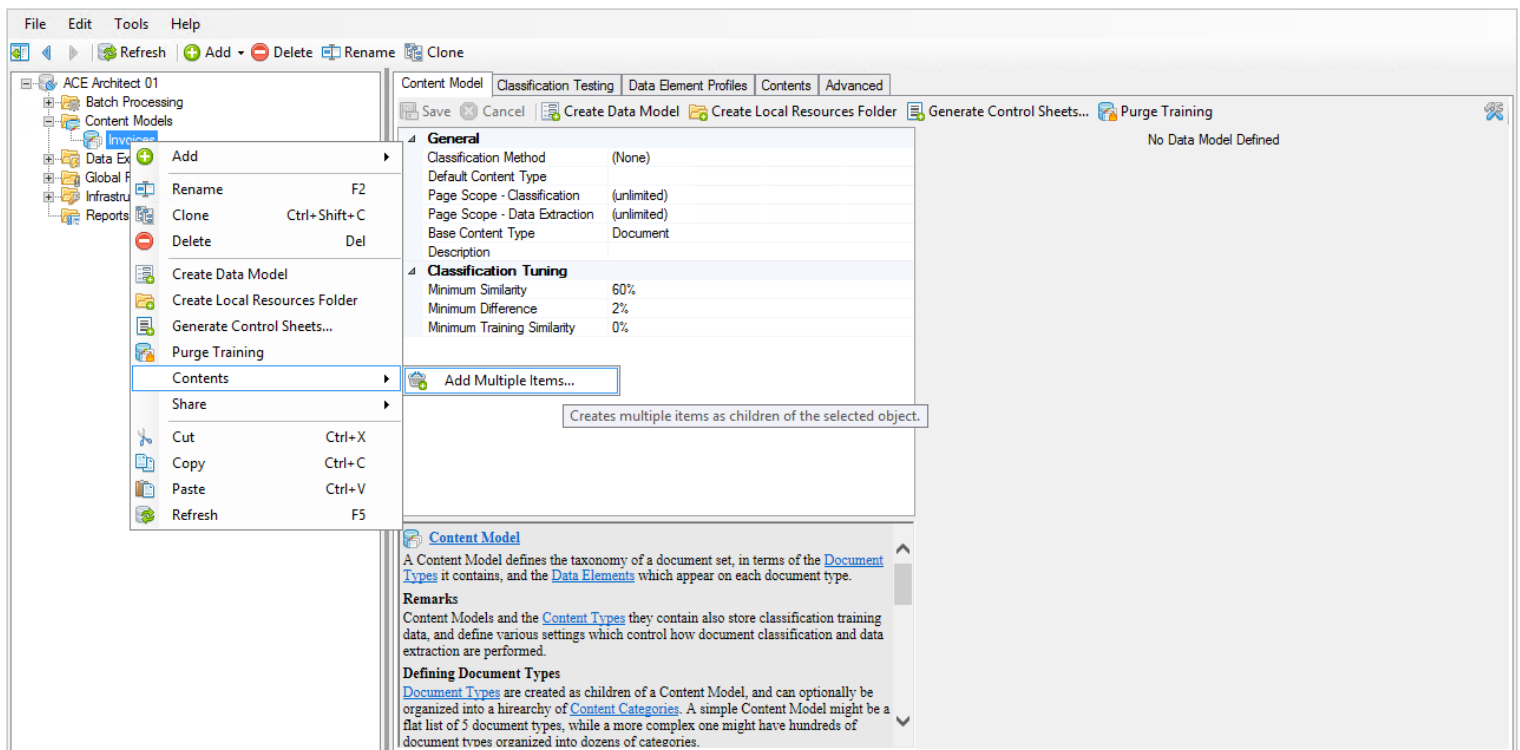
### > Step 1

1. Navigate to (root) > Content Models.
2. Right click on Content Models and select Add > Content Model....
3. Name it **Invoices** .



## > Step 2

1. Right click on the new Invoices Content Model and select Contents > Add Multiple Items....
2. In the Contents • Add Multiple Items window, set Item Type to Document Type
3. Edit the Item Names property and type out the names of each vendor
  - Acme
  - Enid
  - Express
  - Spartan
  - Standard
4. Click Execute to complete





Contents • Add Multiple Items

Execute

Cancel

General

Item Type

Item Names

Document Type

0 entries

Acme  
Enid  
Express  
Spartan  
Standard

Item Names

Type: *String*

The names of the child items to create.

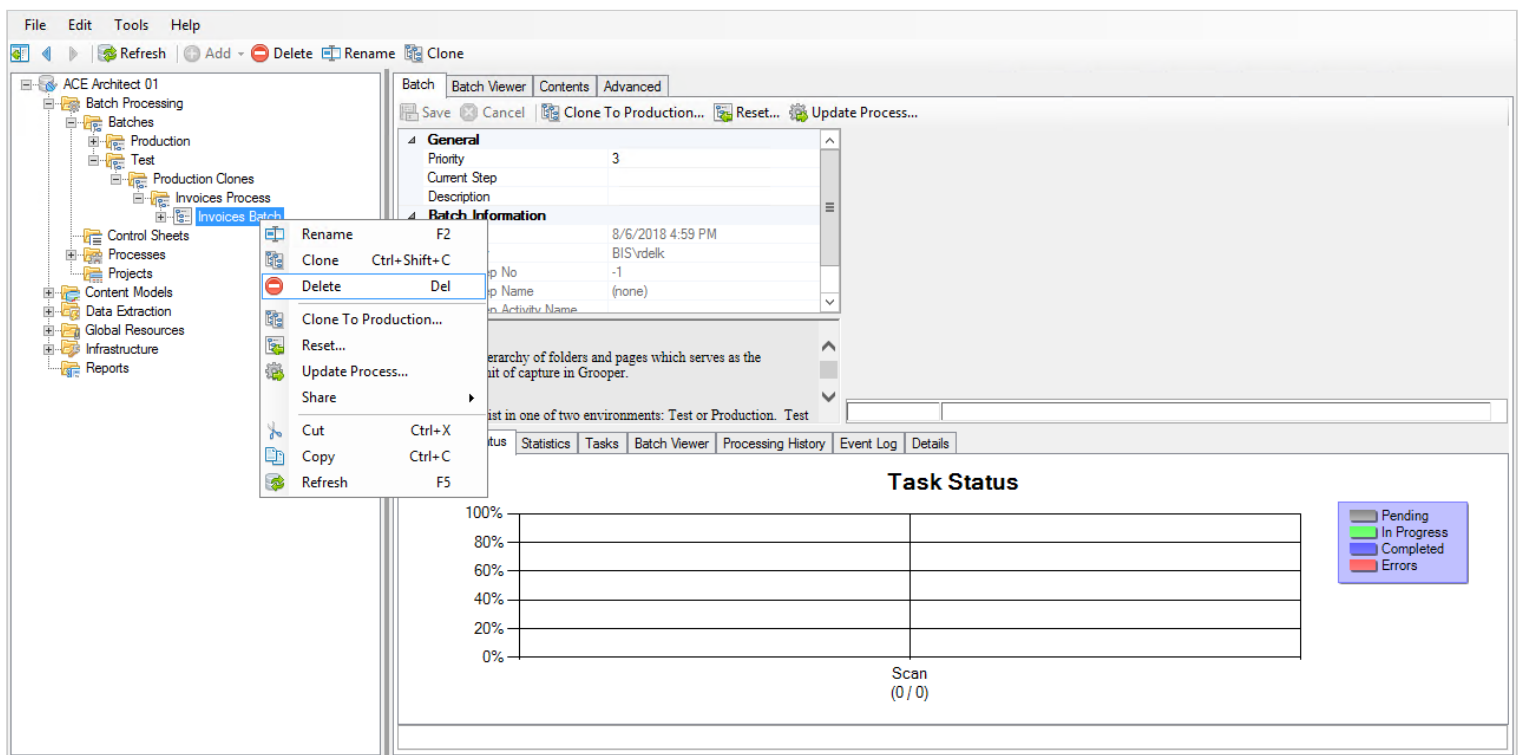
Remarks

For each name provided in the list, a child item of the specified Item Type will be created.

## Re-cloning the production batch

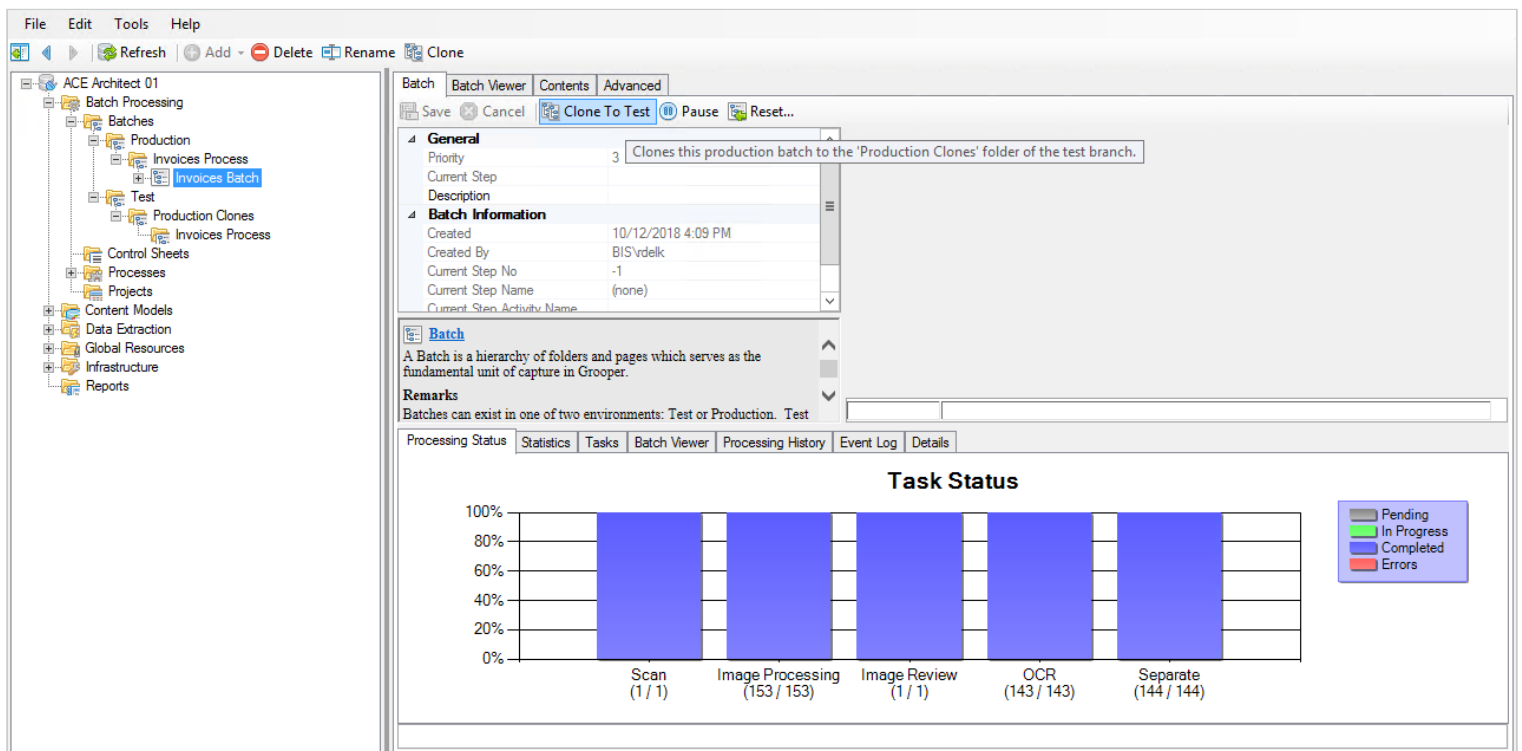
### > Step 1

1. Navigate to (root) > Batch Processing > Batches > Test > Production Clones > Invoices Process > Invoices Batch
2. Right click on Invoices Batch and select Delete



## > Step 2

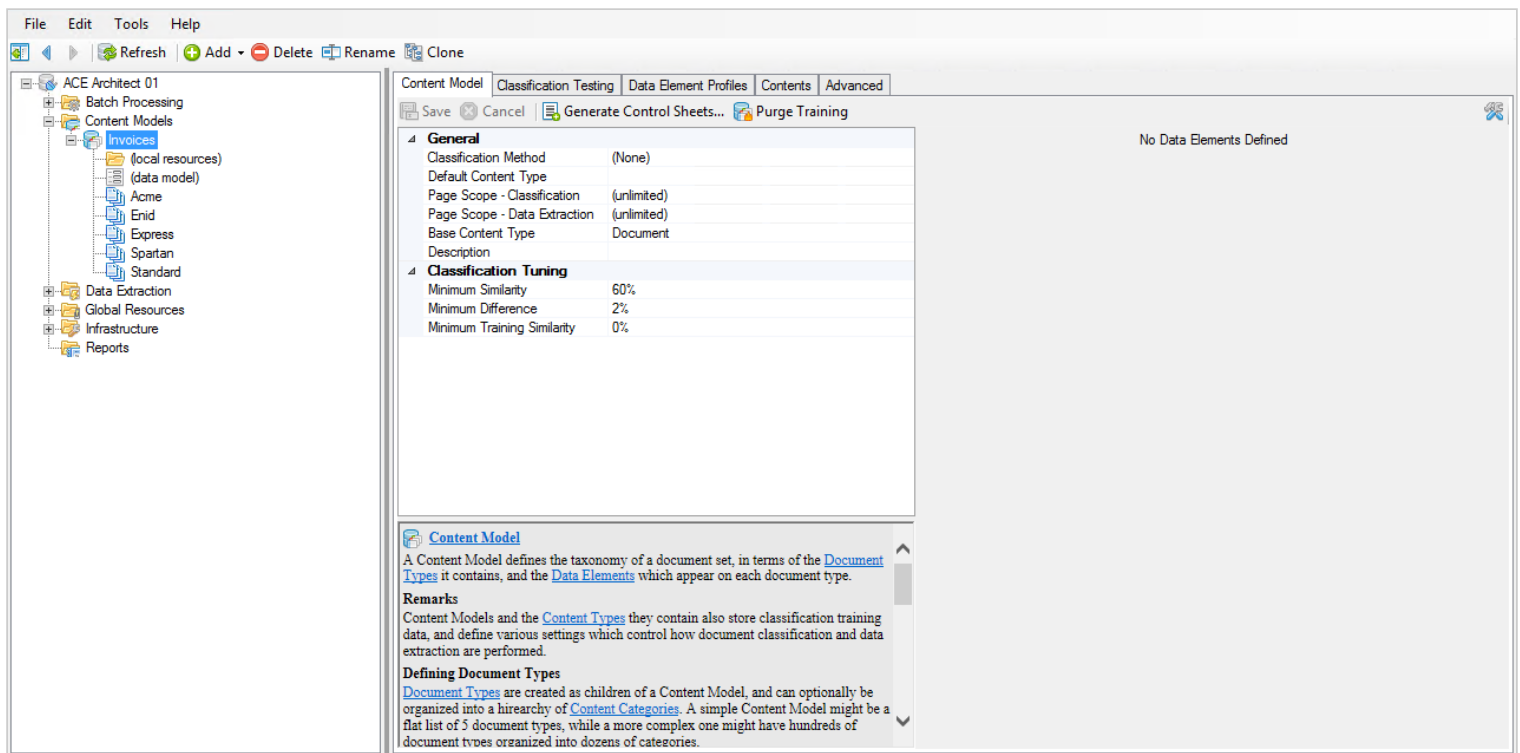
1. Navigate to (root) > Batch Processing > Batches > Production > Invoices Process > Invoices Batch
2. Click Clone To Test



# Setting the classification method

## > Step 1

1. Navigate to (root) > Content Models > Invoices
2. Right click on Invoices and select Create Data Model
3. Right click on Invoices and select Create Local Resources Folder



## > Step 2

1. On the Invoices Content Model in the General section, set the Classification Method property to **Lexical**
2. Still in the General section, navigate to Classification Method > Text Feature Extractor > Type
3. Set the Type property to **Reference**
4. Set the Referenced Extractor property to the Data Extraction > Downloads > Features > **Words (Stemmed)** Data Type

File Edit Tools Help

Refresh Add Delete Rename Clone

ACE Architect 01

- Batch Processing
- Content Models
  - Invoices
    - (local resources)
    - (data model)
    - Acme
    - Enid
    - Express
    - Spartan
    - Standard
  - Data Extraction
  - Global Resources
  - Infrastructure
  - Reports

Content Model Classification Testing Data Element Profiles Contents Advanced

Save Cancel

No Data Elements Defined

**General**

Classification Method: **Lexical**

Default Content Type: (None)

Page Scope - Classification: **Lexical**

Page Scope - Data Extraction: **Visual**

Base Content Type: Visual

Description:

**Classification Tuning**

Minimum Similarity	60%
Minimum Difference	2%
Minimum Training Similarity	0%

Classifies documents based on their text content, using preconfigured training and/or rules. Lexical classification is configured by defining a set of Document Types, and then teaching Grooper to recognize each document type. This can be done by training Grooper with one or more samples of the document, or by defining hand-coded rules which identify the document type. Training-Based ClassificationThe training-based approach measures document similarity by analyzing the frequency of features which appear in the document. In the simplest case, a 'feature' is an individual word, and training is a process of recording the word frequencies of each document type. At classification time, the word frequencies found on a document will be compared to the word frequencies of document types in the training database, generating a similarity value for each document type. Training can be performed using the Content Type - Classification Testing tab, using the Classify Review activity, or using an instance of the Review

**Classification Method**

Type: [Classify Method](#)

Specifies the method to be used for training and classifying documents. Can be one of the following values:

- [Lexical](#) - Classifies documents based on their text content, using preconfigured training and/or rules.
- [Visual](#) - Classifies documents or pages based on their visual appearance.

**Remarks**

If no classification method is specified, all documents will be classified as the Default Content Type.

**Current Value:** [Lexical](#)

File Edit Tools Help

Refresh Add Delete Rename Clone

ACE Architect 01

- Batch Processing
- Content Models
  - Invoices
    - (local resources)
    - (data model)
    - Acme
    - Enid
    - Express
    - Spartan
    - Standard
  - Data Extraction
  - Global Resources
  - Infrastructure
  - Reports

Content Model Classification Testing Data Element Profiles Contents Advanced

Save Cancel Generate Control Sheets... Purge Training Rebuild Training

No Data Elements Defined

**General**

Classification Method: **Lexical**

Text Feature Extractor: **Words (Stemmed)**

Type: **Reference**

Referenced Extractor: **Words (Stemmed)**

Image Feature Extractor: (empty)

EPI Extractor: (empty)

Use Class Frequency: False

Sublinear TF Scaling: False

Smooth IDF: True

Default Content Type: (unlimited)

Page Scope - Classification: (unlimited)

Page Scope - Data Extraction: (unlimited)

Base Content Type: Document

Description:

**Classification Tuning**

Minimum Similarity	60%
Minimum Difference	2%
Minimum Training Similarity	0%

**Referenced Extractor**

Type: [Grooper Node](#)

When Type is set to 'Reference', specifies the external extractor to execute.

**Property Type:** [Grooper Node](#)

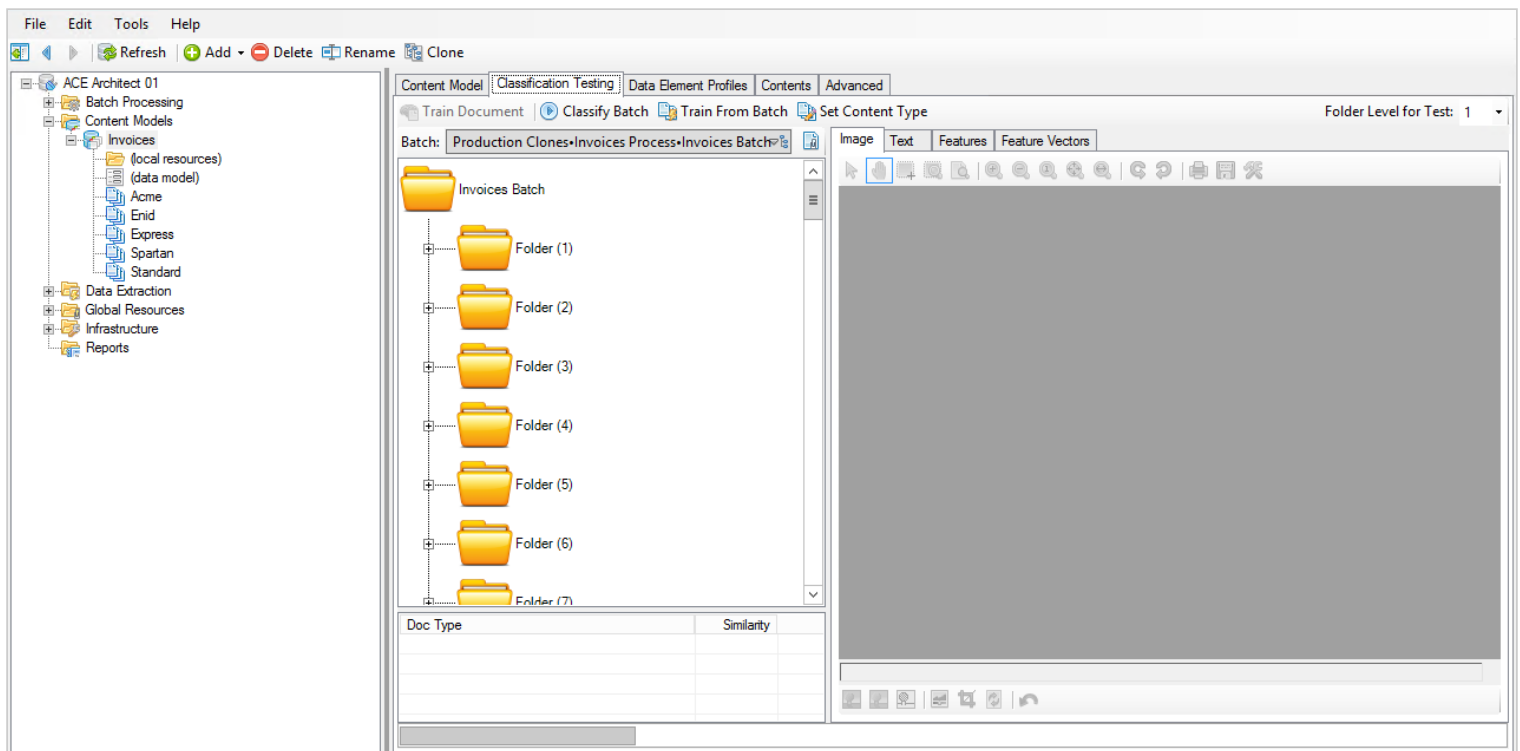
Grooper Nodes are objects which appear in the Grooper Tree Structure. Every object at every level of the tree structure, including folders, are considered Grooper Nodes. All Grooper Node objects support a base set of properties and behaviors common to all node types, such as:

- Name Property** - All node types have a name property.
- Allowed Child Types** - Each node type defines a set of allowed child types, defining what type of objects may be added as children. For example, a [Data Type](#) object may have [Data Types](#) or [Data Formats](#) and children, but cannot

# Training and testing classification

## > Step 1

1. Click on the Classification Testing tab
2. Select Folder (1)
3. In the toolbar, click on the Train Document button
4. Select the Acme Document Type from the list
5. In the toolbar, click on the Classify Batch button



File Edit Tools Help

Refresh Add Delete Rename Clone

ACE Architect 01

- Batch Processing
- Content Models
  - Invoices
    - (local resources)
    - (data model)
    - Acme
    - Enid
    - Express
    - Spartan
    - Standard
  - Data Extraction
  - Global Resources
  - Infrastructure
  - Reports

Content Model Classification Testing Data Element Profiles Contents Advanced

Train Document Classify Batch Train From Batch Set Content Type

Folder Level for Test: 1

Batch: Production Clones-Invoices

Train a Content Type from the contents of the selected document. This does not classify the document.

Invoices Batch

- Folder (1)
- Folder (2)
- Folder (3)
- Folder (4)
- Folder (5)
- Folder (6)
- Folder (7)

Doc Type	Similarity

ACME | INTERNATIONAL

Acme International, Inc  
123 South Main  
Durham, NH 03824  
Phone (603) 333-4444

Checks to:  
Acme International, Inc  
123 South Main  
Durham, NH 03824  
Phone (603) 333-4444

Invoice

Ship To:  
Grooper Industries  
13900 N Harvey  
Edmond, OK 73013  
405-507-7000

Page 1 / 01

Your Reference  
P.O. number : 201017081  
Customer number : 18003405  
Customer contact :  
Currency : USD  
Invoice amount : 1,727.13  
Payment terms : Net 30 days

Our Reference  
Date : 11/14/2008  
Invoice number : 74461405  
Ship-to Number : 18103943  
Sales order no : 33261728  
Delivery number : 6515618  
Freight carrier :  
Tracking number : 1Z2565E90301034108

Item	Material Description	Quantity	Unit Price	Value
000020	GB.C103003-00001	2 EA		
	BRACKET			

Scale: 21 % | 2543px x 3283px | 8.48" x 10.94" | 300 DPI | Black & White

File Edit Tools Help

Refresh Add Delete Rename Clone

ACE Architect 01

- Batch Processing
- Content Models
  - Invoices
    - (local resources)
    - (data model)
    - Acme
    - Enid
    - Express
    - Spartan
    - Standard
  - Data Extraction
  - Global Resources
  - Infrastructure
  - Reports

Content Model Classification Testing Data Element Profiles Contents Advanced

Train Document Classify Batch Train From Batch Set Content Type

Folder Level for Test: 1

Batch: Production Clones-Invoices

Use the training from the current Content Model to classify the documents in the Batch.

Invoices Batch

- Acme (1)
- Folder (2)
- Folder (3)
- Folder (4)
- Folder (5)
- Folder (6)
- Folder (7)

Doc Type	Similarity

ACME | INTERNATIONAL

Acme International, Inc  
123 South Main  
Durham, NH 03824  
Phone (603) 333-4444

Checks to:  
Acme International, Inc  
123 South Main  
Durham, NH 03824  
Phone (603) 333-4444

Invoice

Ship To:  
Grooper Industries  
13900 N Harvey  
Edmond, OK 73013  
405-507-7000

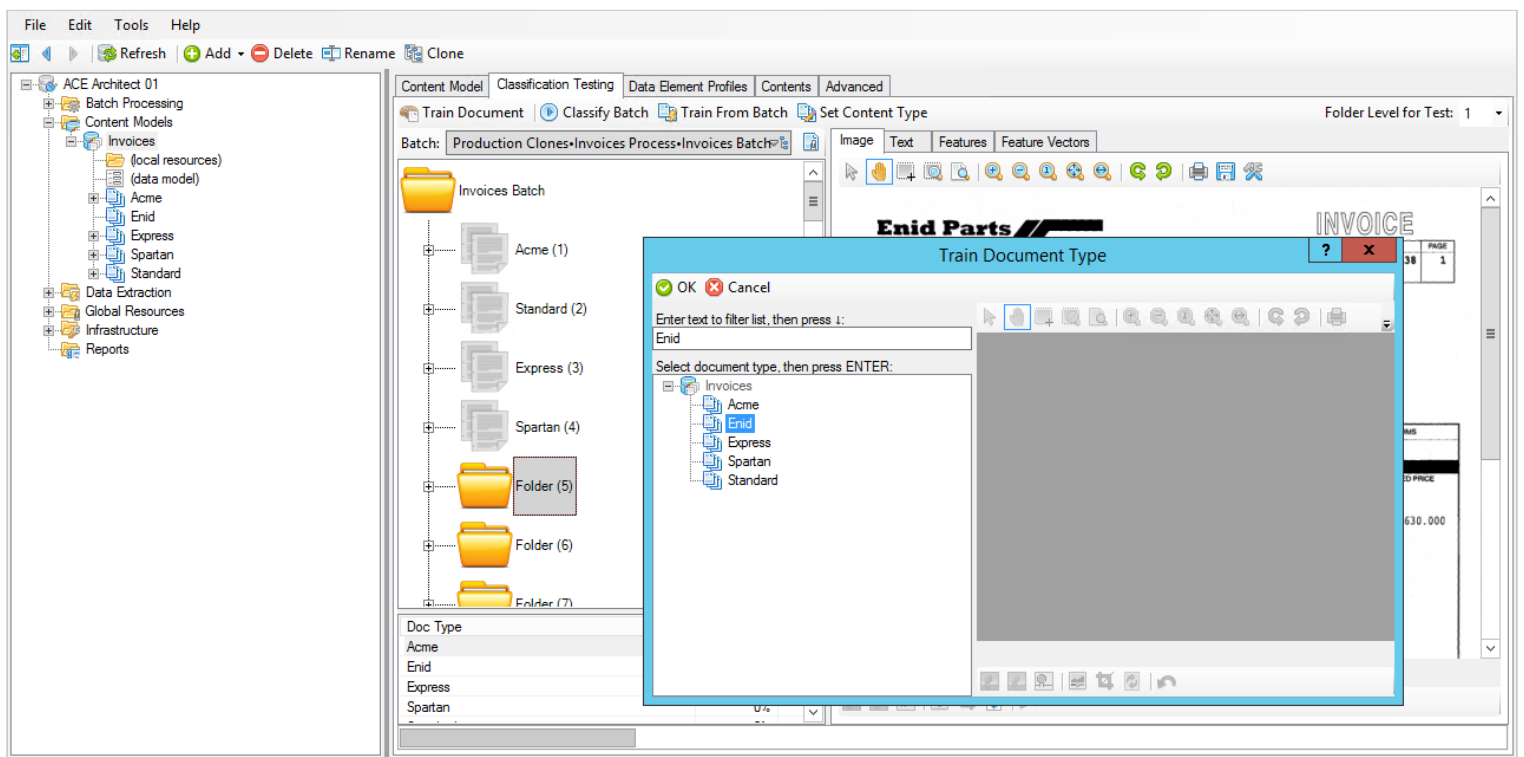
Page 1 / 01

Your Reference  
P.O. number : 201017081  
Customer number : 18003405  
Customer contact :  
Currency : USD  
Invoice amount : 1,727.13  
Payment terms : Net 30 days

Our Reference  
Date : 11/14/2008  
Invoice number : 74461405  
Ship-to Number : 18103943  
Sales order no : 33261728  
Delivery number : 6515618  
Freight carrier :  
Tracking number : 1Z2565E90301034108

Item	Material Description	Quantity	Unit Price	Value
000020	GB.C103003-00001	2 EA		
	BRACKET			

Scale: 21 % | 2543px x 3283px | 8.48" x 10.94" | 300 DPI | Black & White

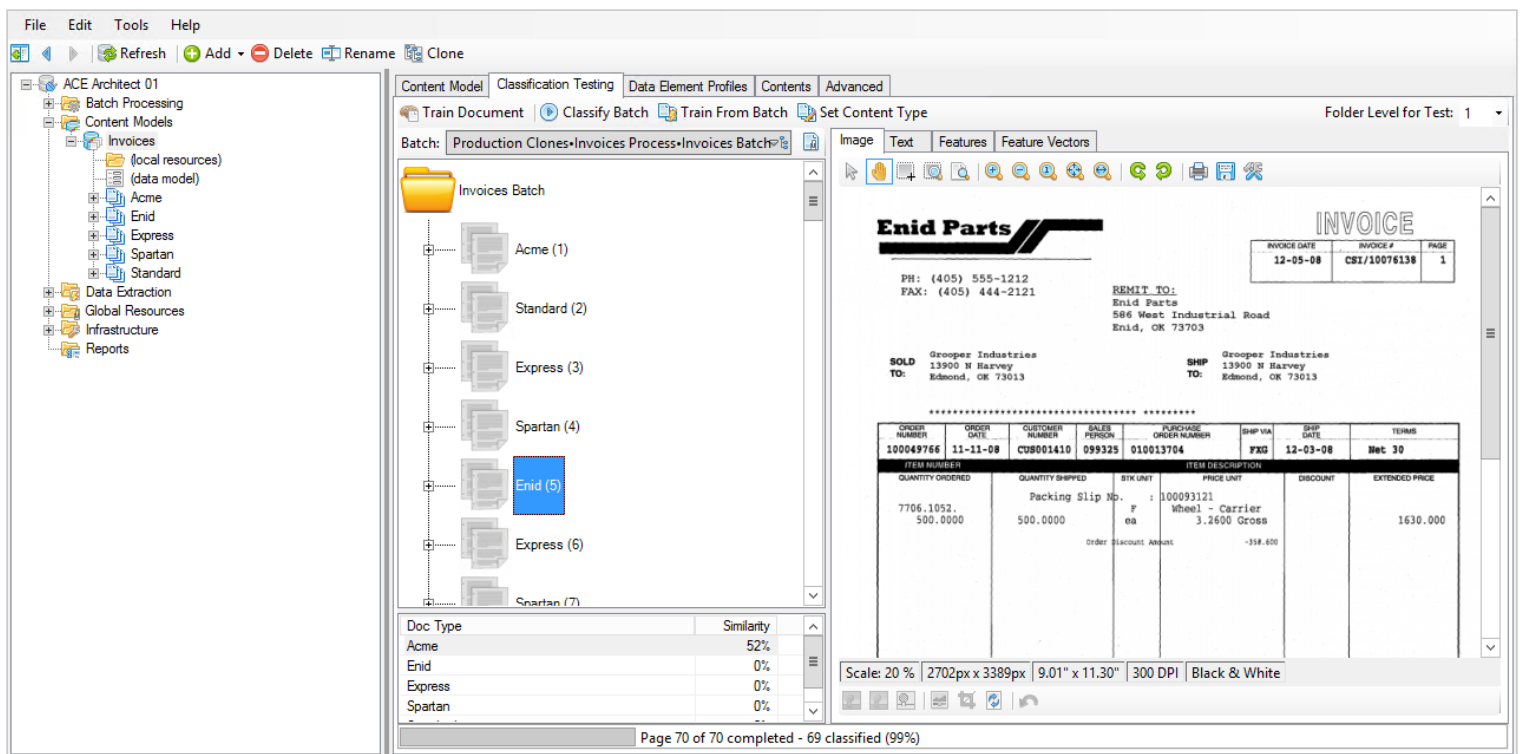
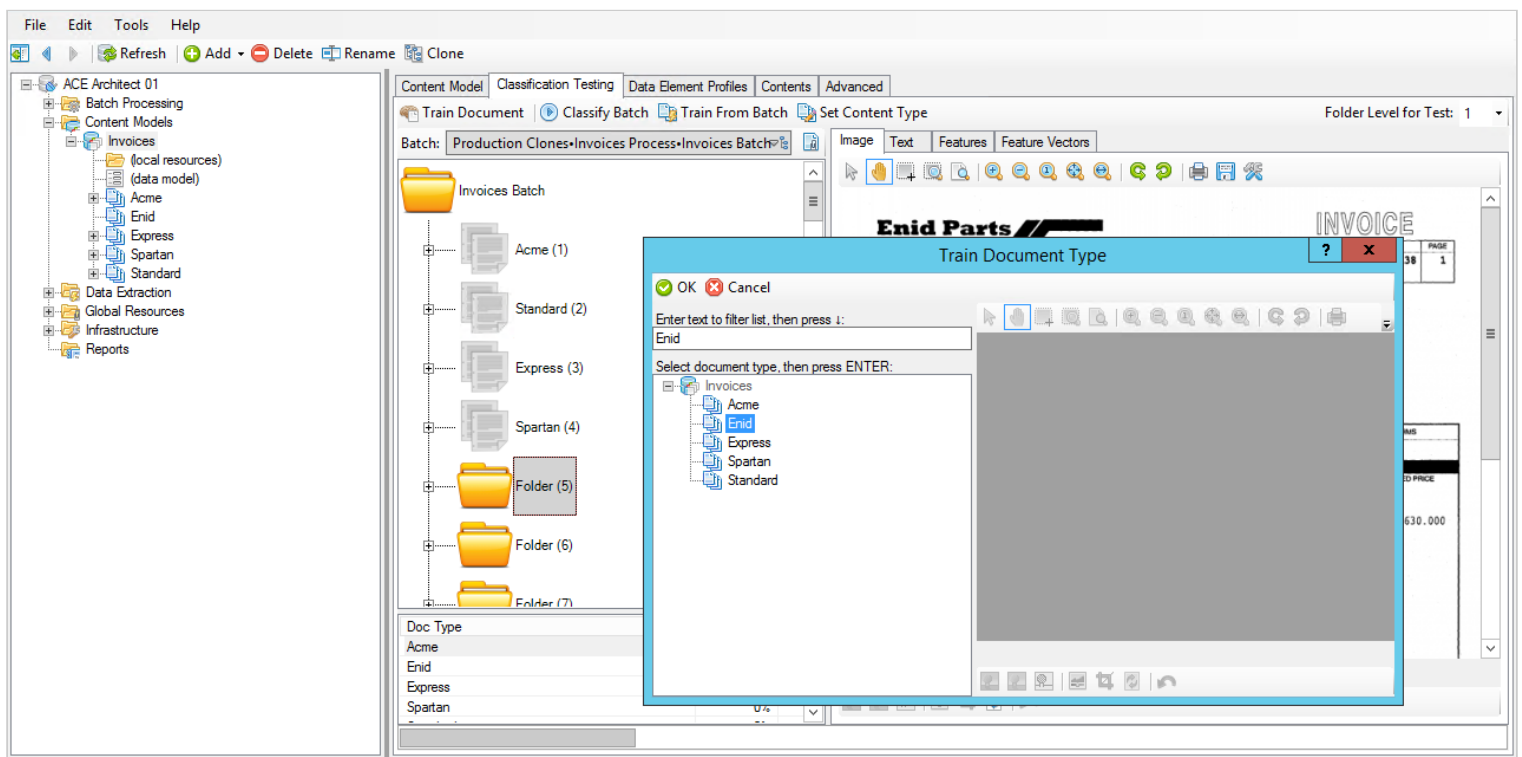


## > Step 2

1. Train the other four vendors:

- Folder (2) Standard (2)
- Folder (3) Express (3)
- Folder (4) Spartan (4)
- Folder (5) Enid (5)

2. In the toolbar, click on the Classify Batch button





# Classify Review

## Adding a Classify Review step

### > Step 1

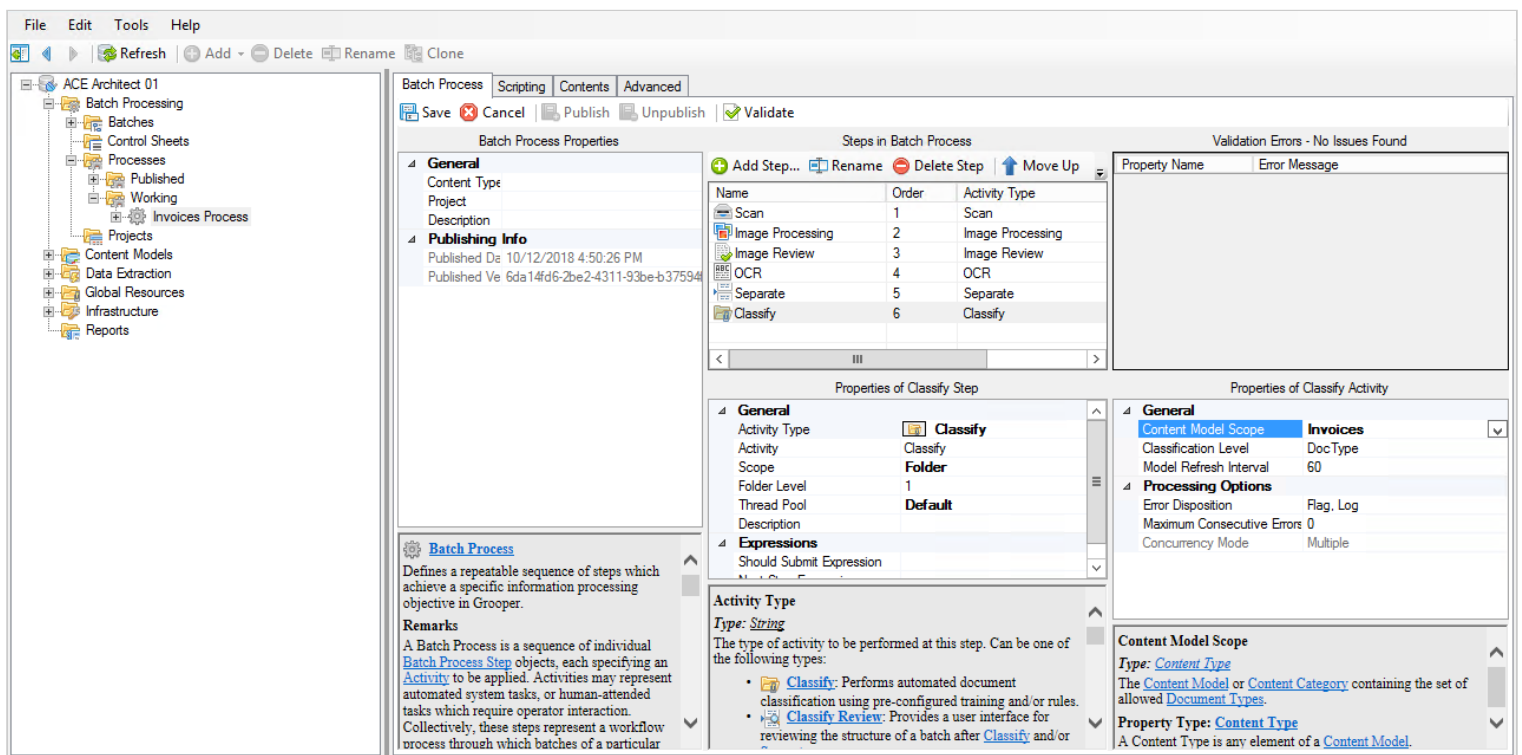
1. Navigate to (root) > Batch Processing > Processes > Working > Invoices Process
2. Add a new step
3. Set the Activity Type property of the new step to Classify
4. Set the Content Model Scope to our Invoices Content Model

The screenshot displays the ACE Architect 01 software interface. On the left, a tree view shows the project structure, with 'Invoices Process' selected under 'Processes > Working'. The main window is divided into several panes:

- Batch Process Properties:** Contains 'General' and 'Publishing Info' tabs. The 'General' tab is active, showing fields for 'Content Type', 'Project', and 'Description'. The 'Publishing Info' tab shows 'Published On' (10/12/2018 4:50:26 PM) and 'Published By' (6da14fd6-2be2-4311-93be-b37594).
- Steps in Batch Process:** A table listing the steps in the process.
- Validation Errors:** A section titled 'Validation Errors - No Issues Found' with a table for 'Property Name' and 'Error Message'.
- Batch Process Description:** A text area at the bottom left containing a definition and remarks about the batch process.

Name	Order	Activity Type
Scan	1	Scan
Image Processing	2	Image Processing
Image Review	3	Image Review
OCR	4	OCR
Separate	5	Separate

**Batch Process**  
Defines a repeatable sequence of steps which achieve a specific information processing objective in Grooper.  
**Remarks**  
A Batch Process is a sequence of individual [Batch Process Step](#) objects, each specifying an [Activity](#) to be applied. Activities may represent automated system tasks, or human-attended tasks which require operator interaction. Collectively, these steps represent a workflow process through which batches of a particular



## > Step 2

1. Add a new step
2. Set the Activity Type property of the new step to Classify Review
3. Set the Classification Viewer Settings > Content Model Scope to our Invoices Content Model
4. Save and Publish.

File Edit Tools Help

Refresh Add Delete Rename Clone

ACE Architect 01

- Batch Processing
  - Batches
  - Control Sheets
  - Processes
    - Published
    - Working
  - Invoices Process
- Projects
- Content Models
- Data Extraction
- Global Resources
- Infrastructure
- Reports

Batch Process Scripting Contents Advanced

Save Cancel Publish Unpublish Validate

Batch Process Properties

General

Content Type

Project

Description

Publishing Info

Published Da 10/12/2018 4:50:26 PM

Published Ve 6da14fd6-2be2-4311-93be-b37594

Steps in Batch Process

Name	Order	Activity Type
Scan	1	Scan
Image Processing	2	Image Processing
Image Review	3	Image Review
OCR	4	OCR
Separate	5	Separate
Classify	6	Classify
Classify Review	7	Classify Review

Validation Errors - 1 Issue Found

Property Name	Error Message
Classify Review • Ac...	Value is required.

Properties of Classify Review Step

General

Activity Type **Classify Review**

Activity Classify Review

Scope **Batch**

Description

Expressions

Should Submit Expression

Next Step Expression

Properties of Classify Review Activity

General

Classification Viewer Settings (Click to edit)

Allow Flagged Folders False

Allow Unclassified Folders False

UI Configuration

Command Options (0 Command Options objects)

User Activity Timeout 0

Classification Viewer Settings

Type: [Classification Viewer Settings](#)

Settings for this control. This optional class allows [Attended Client](#) activities to expose these settings at the [Batch Process Step](#) during configuration in [Grooper Design Studio](#). The settings class must inherit from [BatchView.ControlSettings](#).

Batch Process

Defines a repeatable sequence of steps which achieve a specific information processing objective in Grooper.

Remarks

A Batch Process is a sequence of individual [Batch Process Step](#) objects, each specifying an [Activity](#) to be applied. Activities may represent automated system tasks, or human-attended tasks which require operator interaction. Collectively, these steps represent a workflow process through which batches of a particular

File Edit Tools Help

Refresh Add Delete Rename Clone

ACE Architect 01

- Batch Processing
  - Batches
  - Control Sheets
  - Processes
    - Published
    - Working
  - Invoices Process
- Projects
- Content Models
- Data Extraction
- Global Resources
- Infrastructure
- Reports

Batch Process Scripting Contents Advanced

Save Cancel Publish Unpublish Validate

Classification Viewer Settings

OK Cancel

General

Content Model Scope **Invoices**

Classification Level

DocType

Allow Training False

Flag Messages (empty)

Content Model Scope

Type: [Content Type](#)

The [Content Model](#) or [Content Category](#) containing the set of allowed [Content Types](#).

Property Type: [Content Type](#)

A Content Type is any element of a [Content Model](#).

Validation Errors - 1 Issue Found

Property Name	Error Message
Classify Review • Ac...	Value is required.

Properties of Classify Review Activity

General

Classification Viewer Settings (Click to edit)

Allow Flagged Folders False

Allow Unclassified Folders False

UI Configuration

Command Options (0 Command Options objects)

User Activity Timeout 0

Classification Viewer Settings

Type: [Classification Viewer Settings](#)

Settings for this control. This optional class allows [Attended Client](#) activities to expose these settings at the [Batch Process Step](#) during configuration in [Grooper Design Studio](#). The settings class must inherit from [BatchView.ControlSettings](#).

Batch Process

Defines a repeatable sequence of steps which achieve a specific information processing objective in Grooper.

Remarks

A Batch Process is a sequence of individual [Batch Process Step](#) objects, each specifying an [Activity](#) to be applied. Activities may represent automated system tasks, or human-attended tasks which require operator interaction. Collectively, these steps represent a workflow process through which batches of a particular

# Updating the production batch

## > Step 1

1. Navigate to (root) > Batch Processing > Batches > Production > Invoices Process
2. Make sure the batch is paused
3. Right click on the Invoices Batch and select Update Process....
4. Set Target Step to Classify.
5. Click Execute.

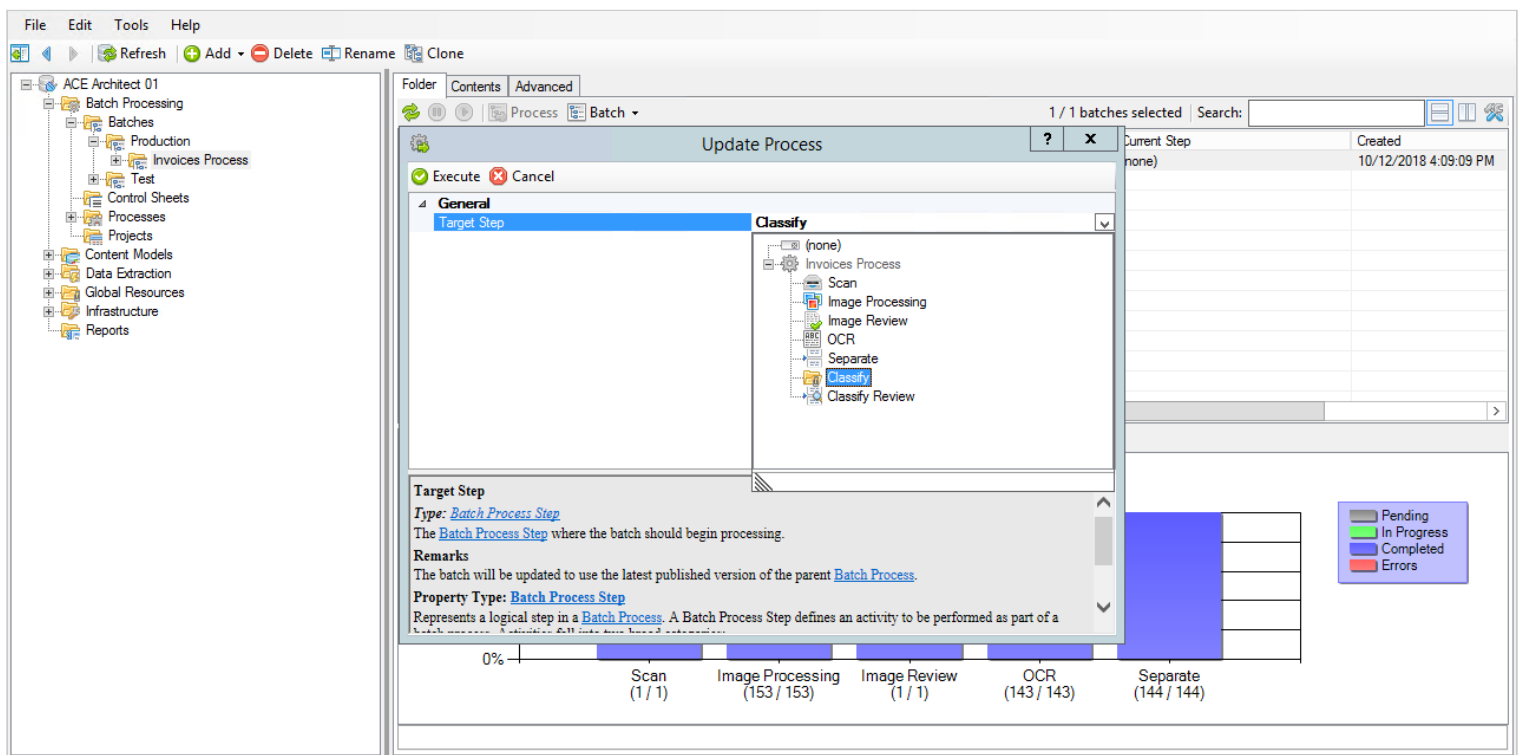
The screenshot displays the ACE Architect 01 software interface. On the left, a tree view shows the project structure under 'ACE Architect 01', with 'Batch Processing' > 'Batches' > 'Production' > 'Invoices Process' selected. The main window shows a table of batch processes. The 'Invoices Batch' is selected, and a context menu is open with 'Update Process...' highlighted. A tooltip for 'Update Process...' reads: 'Updates or changes the Batch Process associated with this batch.'

Batch Name	Priority	Status	Batch Process	Current Step	Created
Invoices Batch	3	Paused	Invoices Process	(none)	10/12/2018 4:09:09 PM

Below the table, the 'Task Status' section shows a bar chart with the following data:

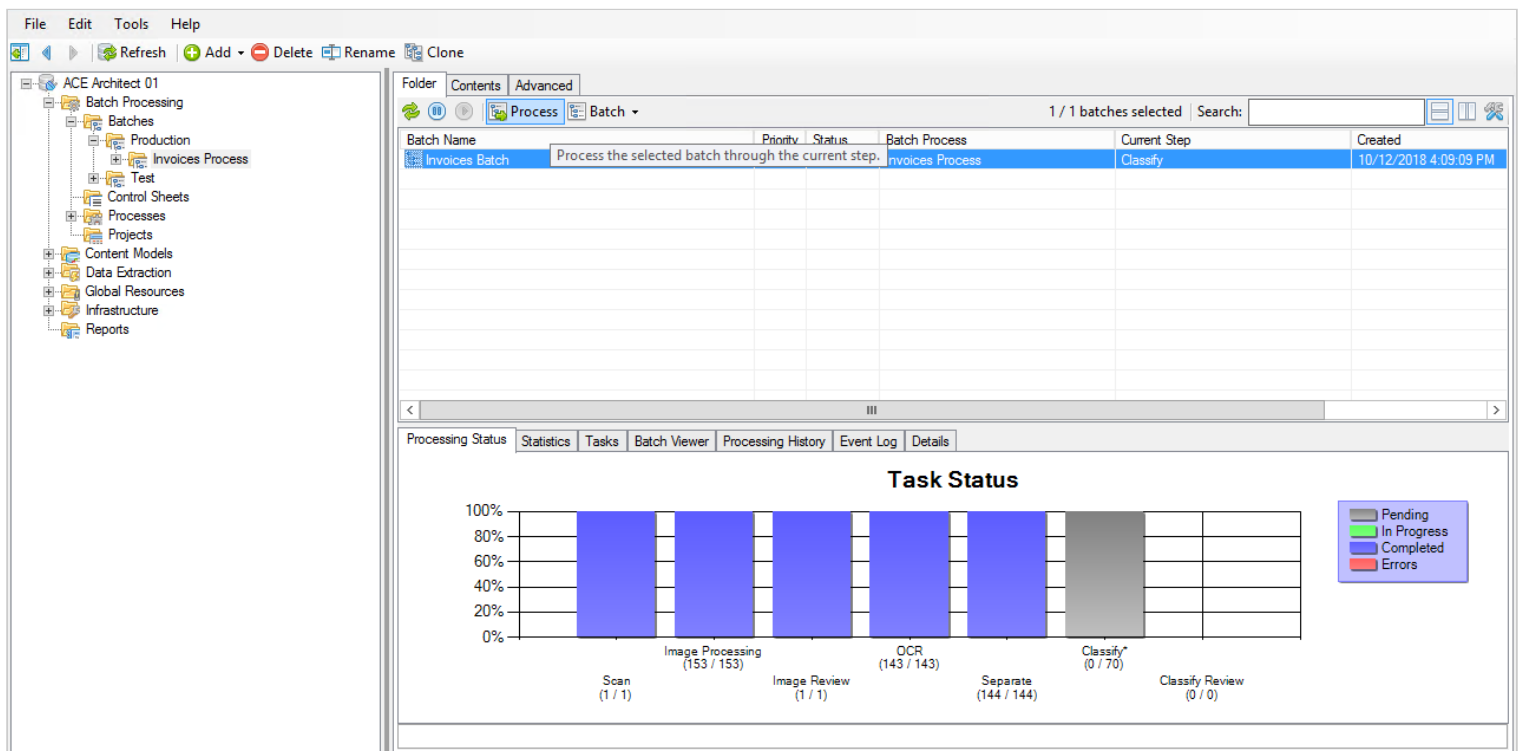
Task	Count	Status
Scan	1 / 1	Completed
Image Processing	153 / 153	Completed
Image Review	1 / 1	Completed
OCR	143 / 143	Completed
Separate	144 / 144	Completed

The legend indicates: Pending (grey), In Progress (green), Completed (blue), and Errors (red). All tasks shown are in the 'Completed' state.



## > Step 2

1. Resume the batch
2. Click Process



# Reviewing and correcting classification

## > Step 1

- 1. After the Classify step has completed, click Process to start the Classify Review step.
- 2. In Grooper Classify Review, right click on Folder (18) and select Assign Content Type
- 3. Select the Enid Document Type from the list and click OK
- 4. In the toolbar, click on the Complete Task button to finish this activity

ProcessingTaskBatchHelp

StartStopOpen Task...Close TaskComplete TaskNew Batch...Settings...

Invoices Batch

Acme (1)

Standard (2)

Express (3)

Spartan (4)

Enid (5)

Document Type Candidates	%
Acme	100%
Enid	28%
Express	20%
Standard	19%
Spartan	17%

General

Batch Process NameInvoices ProcessBatch NameInvoices BatchBatch Open Time00:00:35

ACME | INTERNATIONAL

Acme International, Inc  
123 South Main  
Durham, NH 03824  
Phone (603) 333-4444

Checks to:  
Acme International, Inc  
123 South Main  
Durham, NH 03824  
Phone (603) 333-4444

Invoice

Ship To:  
Grooper Industries  
13900 N Harvey  
Edmond, OK 73013  
405-507-7000

Page  
1 / 01

Bill To:  
Grooper Industries  
13900 N Harvey  
Edmond, OK 73013  
405-507-7000

Our Reference

P.O. number : 201017081Customer number : 18003405Customer contact :Currency : USDInvoice amount : 1,727.13Payment terms : Net 30 days

Our Reference

Date : 11/14/2008Invoice number : 74451405Ship-to Number : 18103943Sales order no : 33241729Delivery number : 5515616Freight carrier :Tracking number : 1Z2565E90301034108

Scale: 34 % | 2543px x 3283px | 8.48" x 10.94" | 300 DPI | Black & White

Processing Task Batch Help

Start Stop Open Task... Close Task Complete Task New Batch... Settings...

Standard (15)

Acme (16)

Express (17)

Folder (18)

Acme

Express

Cut Ctrl+X

Copy Ctrl+C

Paste Ctrl+V

Delete Del

Append to Previous Ctrl+P

Assign Content Type Ctrl+F

Flag Item... Assigns a Content Type to the currently selected folder.

Go To Next Issue Ctrl+I

Go To Next Flagged Item Ctrl+G

New Folder Ins

Unclassify Ctrl+U

Batch Operations

Contents

Send To

Move Up Ctrl+Up

Move Down Ctrl+Down

Item Properties...

Document Type Candidates

Enid

Acme

Express

Spartan

Standard

General

Batch Process Name

Batch Name

Batch Open Time

Enid Parts

PH: (405) 555-1212

FAX: (405) 444-2121

REMIT TO:

Enid Parts

586 West Industrial Road

Enid, OK 73703

SOLD Grooper Industries

TO: 13900 N Harvey

Edmond, OK 73013

SHIP Grooper Industries

TO: 13900 N Harvey

Edmond, OK 73013

INVOICE

INVOICE DATE	INVOICE #	PAGE
12-08-08	CSI/10076269	1

ORDER NUMBER	ORDER DATE	CUSTOMER NUMBER	SALES PERSON	PURCHASE ORDER NUMBER	SHIP VIA	SHIP DATE	TERMS
100050218	12-04-08	CUS001605	099321	201017541	FXG	12-05-08	Net 30

ITEM NUMBER	QUANTITY ORDERED	QUANTITY SHIPPED	STK UNIT	PRICE UNIT	DISCOUNT	EXTENDED PRICE
7710.1111.	5.0000	4.0000	E	Housing, Cold Head		2877.120
7710.1104.	4.0000	5.0000	ea	719.2800 Gross		
7710.1140.			G	Arm, Latch Cold Head		
			ea	194.4000 Gross		972.000
			E	Cover Rear Hot Head		

Scale: 32 % 2750px x 3437px 9.17" x 11.46" 300 DPI Black & White

Processing Task Batch Help

Start Stop Open Task... Close Task Complete Task New Batch... Settings...

Standard (15)

Acme (16)

Express (17)

Enid (18)

Acme (19)

Express (20)

Document Type Candidates

	%
Enid	58%
Acme	20%
Express	20%
Spartan	20%
Standard	17%

General

Batch Process Name

Batch Name

Batch Open Time

Assign Content Type

Enid Parts

PH: (405) 555-1212

FAX: (405) 444-2121

REMIT TO:

Enid Parts

586 West Industrial Road

Enid, OK 73703

SOLD Grooper Industries

TO: 13900 N Harvey

Edmond, OK 73013

SHIP Grooper Industries

TO: 13900 N Harvey

Edmond, OK 73013

INVOICE

INVOICE DATE	INVOICE #	PAGE
12-08-08	CSI/10076269	1

ORDER NUMBER	ORDER DATE	CUSTOMER NUMBER	SALES PERSON	PURCHASE ORDER NUMBER	SHIP VIA	SHIP DATE	TERMS
100050218	12-04-08	CUS001605	099321	201017541	FXG	12-05-08	Net 30

ITEM NUMBER	QUANTITY ORDERED	QUANTITY SHIPPED	STK UNIT	PRICE UNIT	DISCOUNT	EXTENDED PRICE
7710.1111.	5.0000	4.0000	E	Housing, Cold Head		2877.120
7710.1104.	4.0000	5.0000	ea	719.2800 Gross		
7710.1140.			G	Arm, Latch Cold Head		
			ea	194.4000 Gross		972.000
			E	Cover Rear Hot Head		

Scale: 32 % 2750px x 3437px 9.17" x 11.46" 300 DPI Black & White

