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:

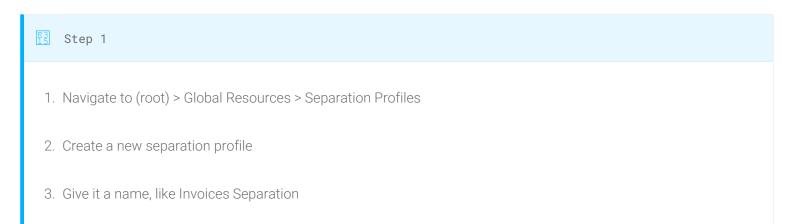
- · create a Separation Profile,
- · use the new profile to separate our invoices into individual folders,
- learn about classification and content types,
- · train Grooper to recognize our documents so it knows what information to get from them, and
- update our Batch Process to do all of this automatically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

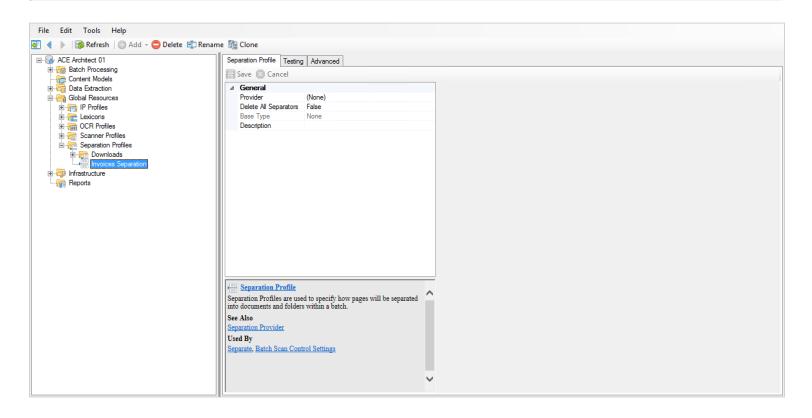
Let's get organized!

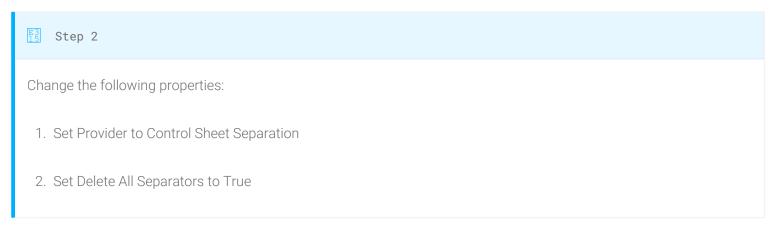
Separation

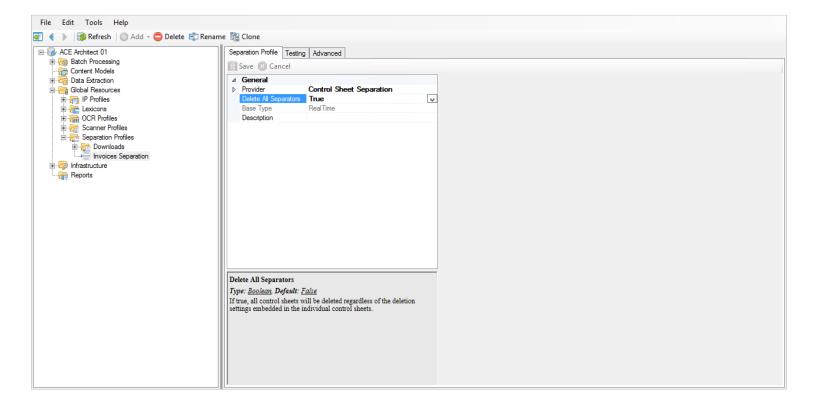
Before we can expect to get any kind of information from these invoices, we need to sort them. While we could get away with doing this manually, Grooper most certainly provides us with a better way.

Creating a Separation Profile

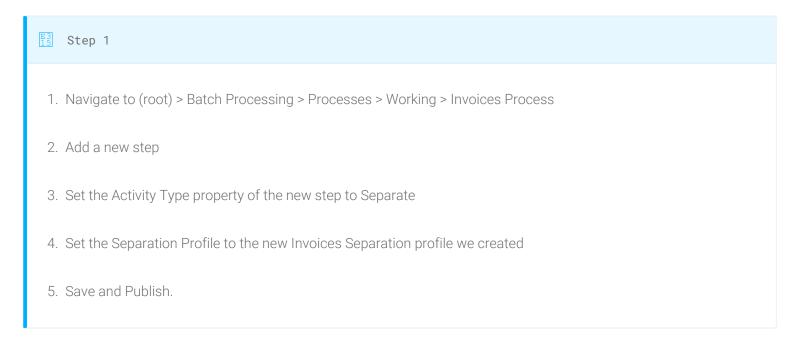


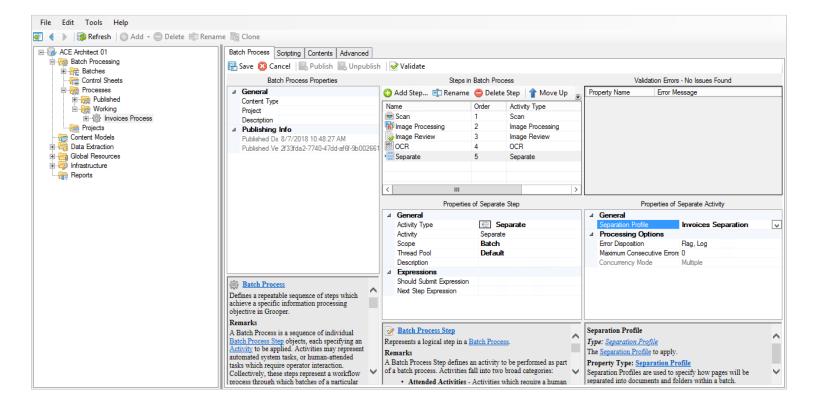






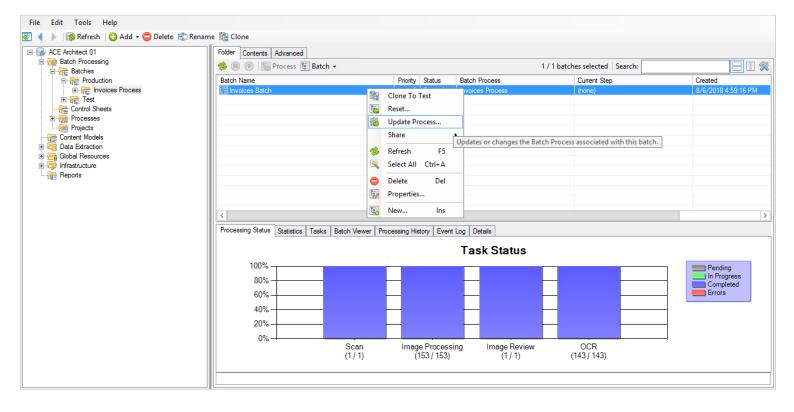
Updating the production batch

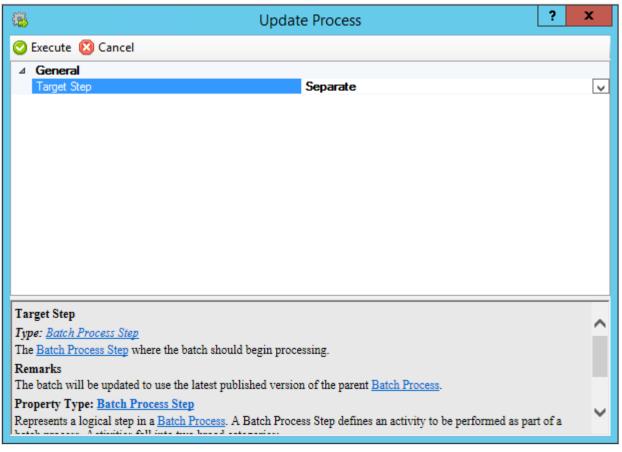


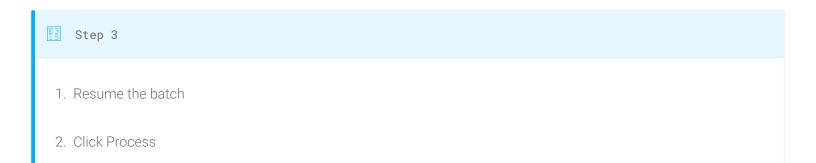


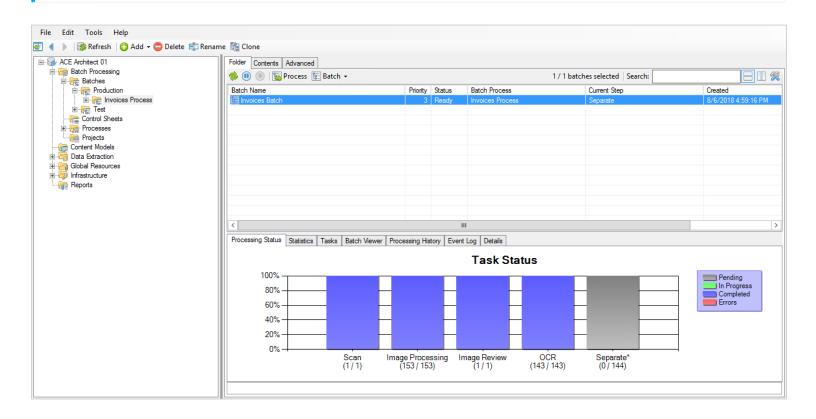


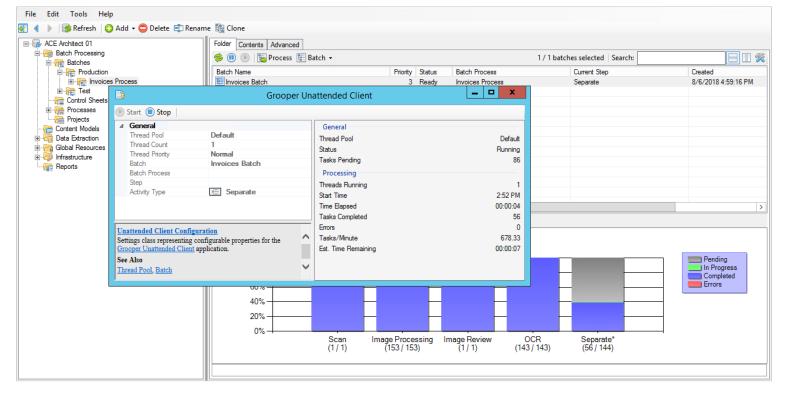
- 1. Navigate to (root) > Batch Processing > Batches > Production > Invoices Process.
- 2. Make sure the batch is paused
- 3. Right click on the Invoices Batch and select Update Process....
- 4. Set Target Step to Separate.
- 5. Click Execute.





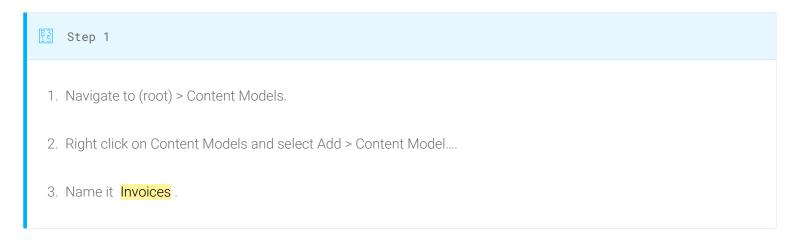


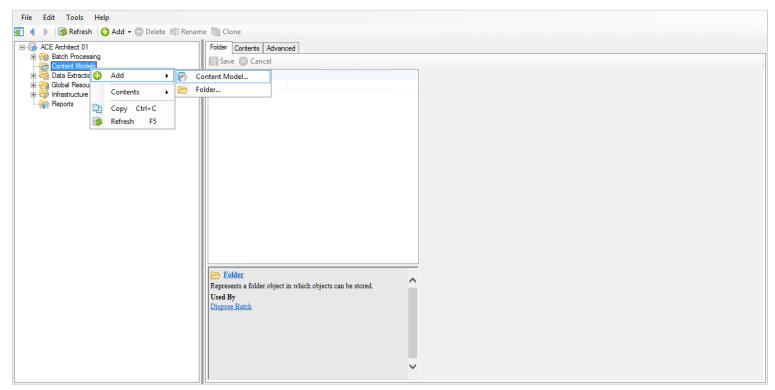




Classification

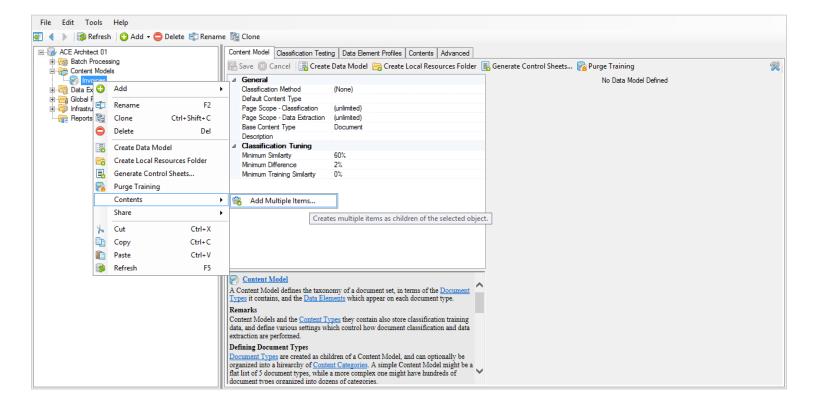
Setting up the Content Model

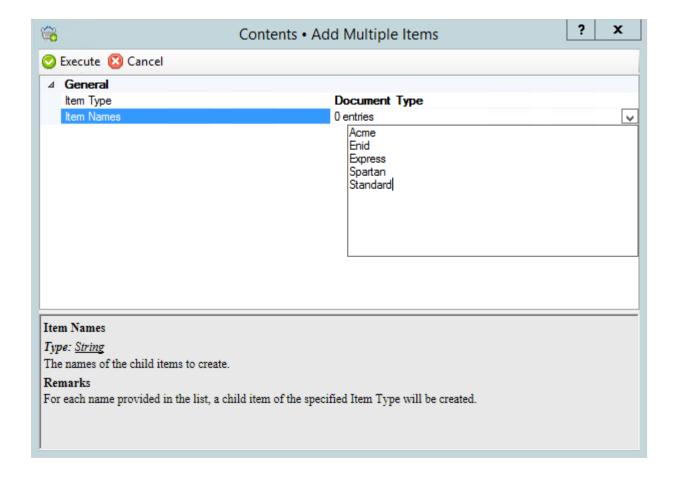




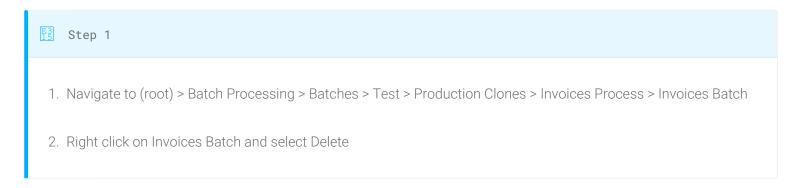


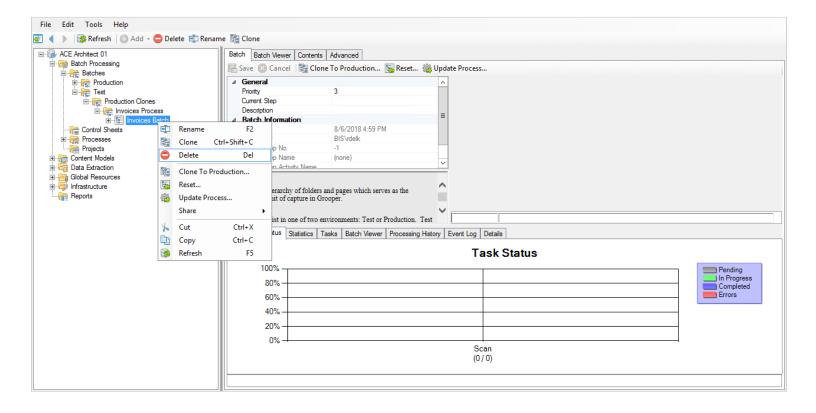
- 1. Right click on the new Invoices Content Model and select Contents > Add Multiple Items....
- 2. In the Contents Add Multiple Items window, set Item Type to Document Type
- 3. Edit the Item Names property and type out the names of each vendor
 - Acme
 - Enid
 - Express
 - Spartan
 - Standard
- 4.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





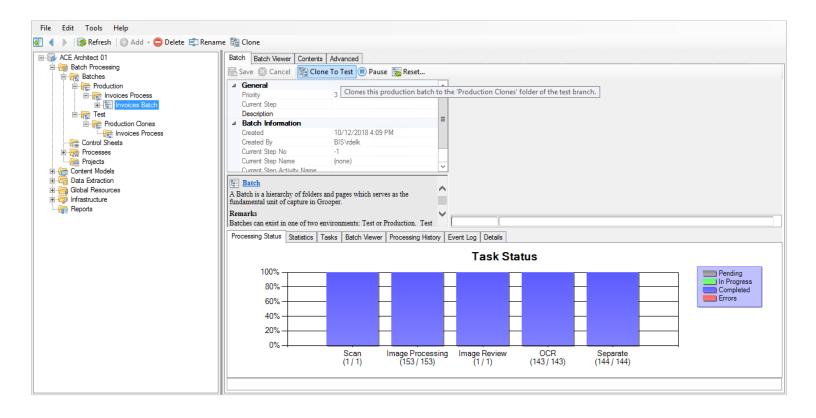
Re-cloning the production batch







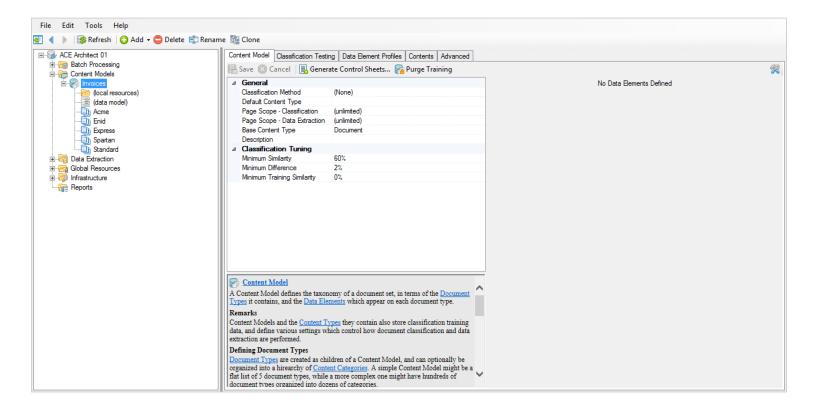
- 1. Navigate to (root) > Batch Processing > Batches > Production > Invoices Process > Invoices Batch
- 2. Click Clone To Test



Setting the classification method

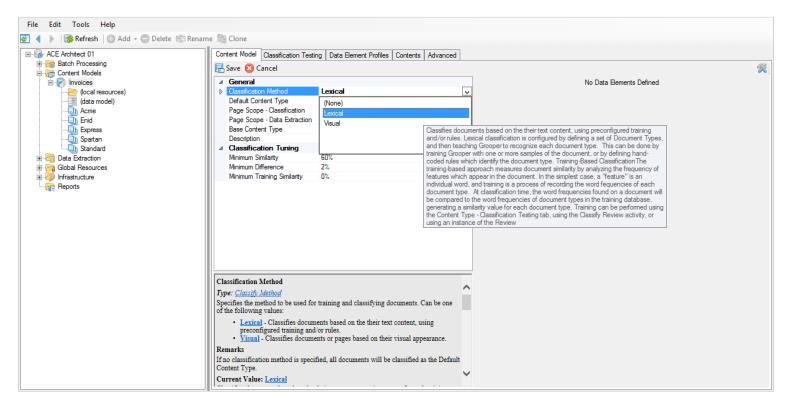


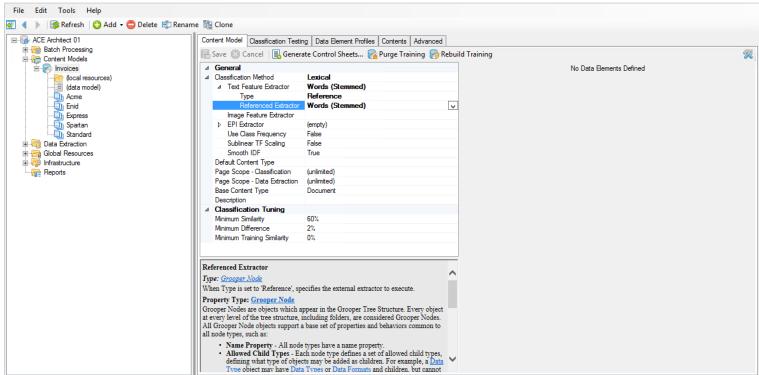
- 1. Navigate to (root) > Content Models > Invoices
- 2. Right click on Invoices and select Create Data Model
- 3. Right click on Invoices and select Create Local Resources Folder



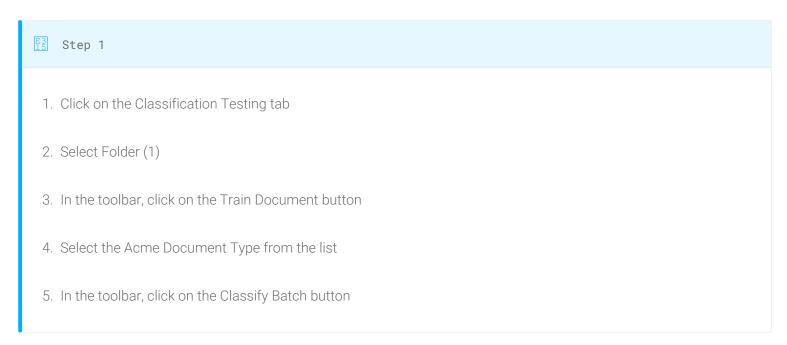


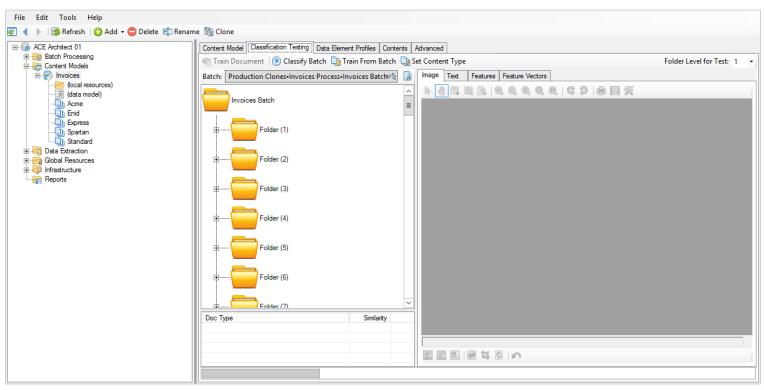
- 1. On the Invoices Content Model in the General section, set the Classification Method property to Lexical
- 2. Still in the General section, navigate to Classification Method > Text Feature Extractor > Type
- 3. Set the Type property to Reference
- 4. Set the Referenced Extractor property to the Data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

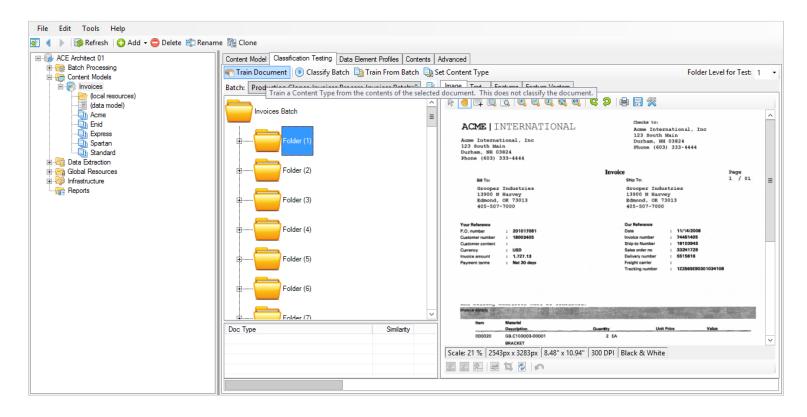


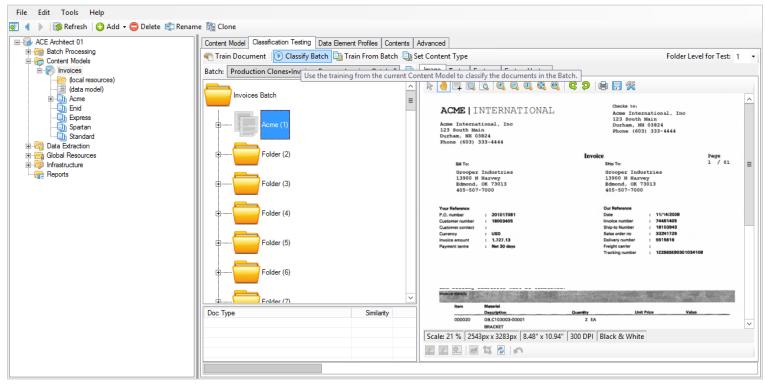


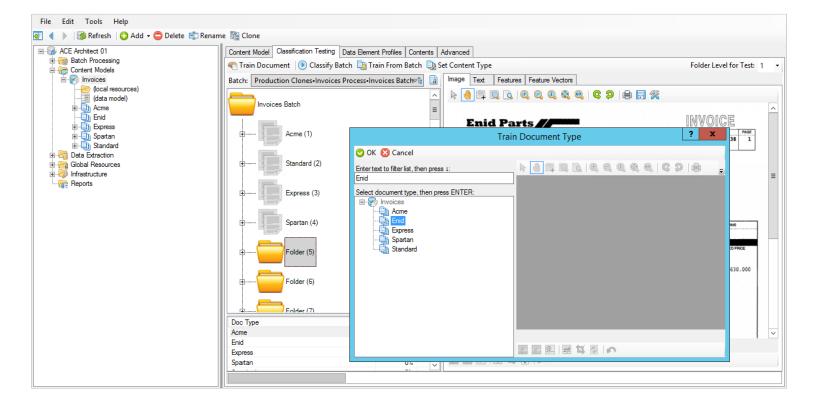
Training and testing classification







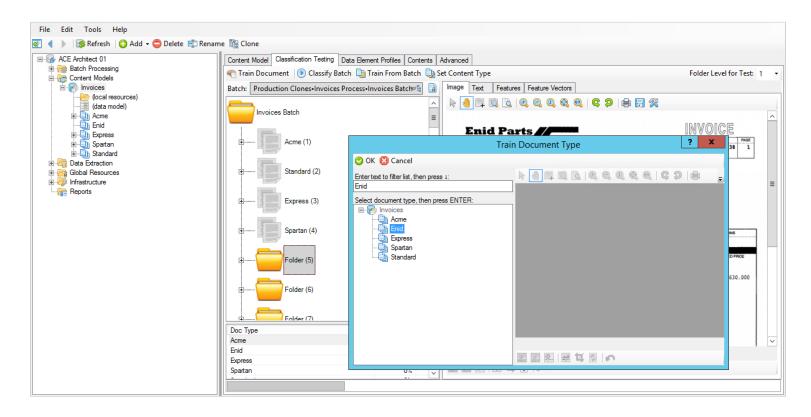


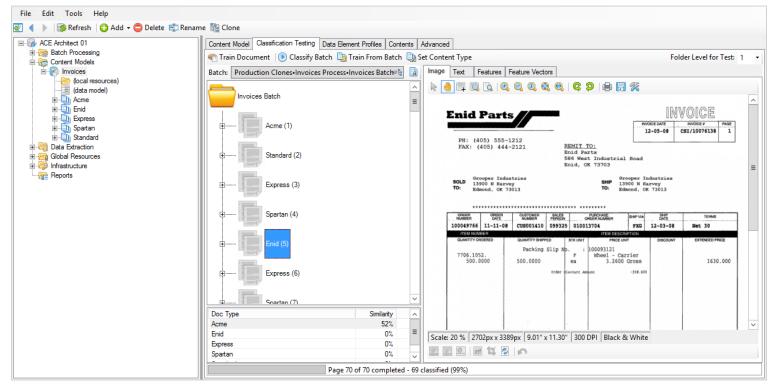




• Folder (5) 🖾 Enid (5)

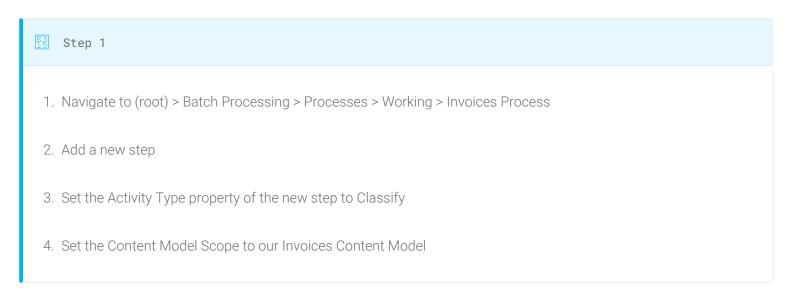
2. In the toolbar, click on the Classify Batch button

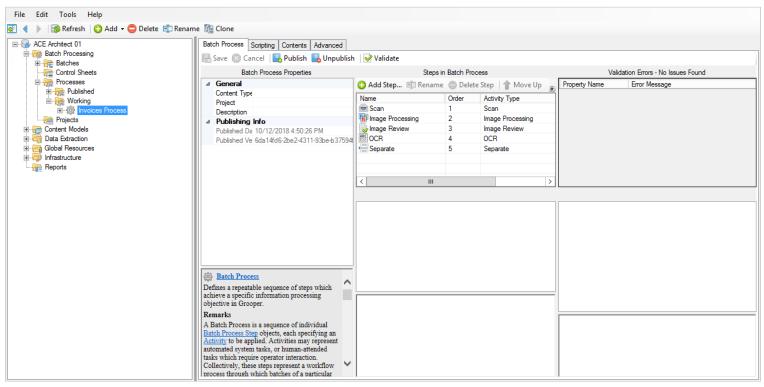


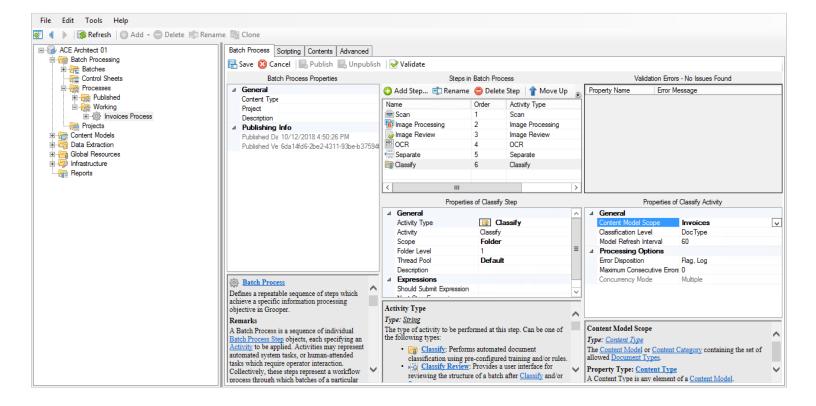


Classify Review

Adding a Classify Review step

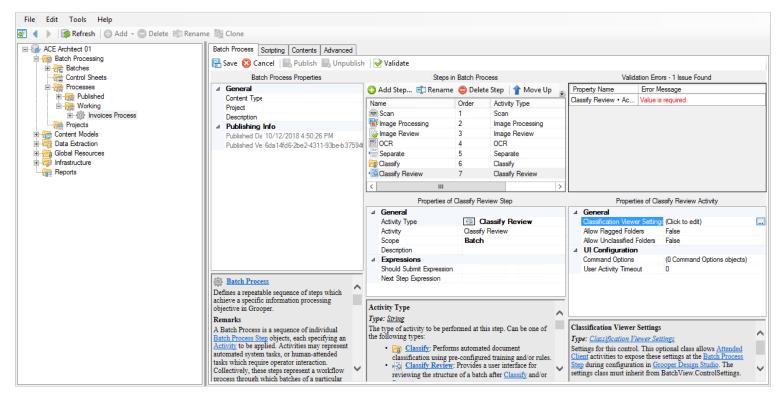


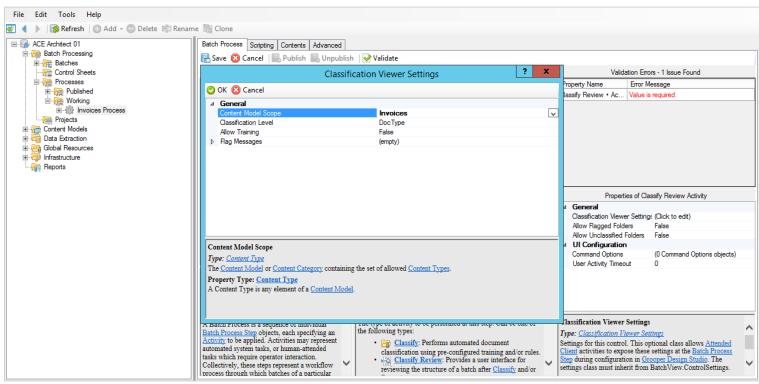




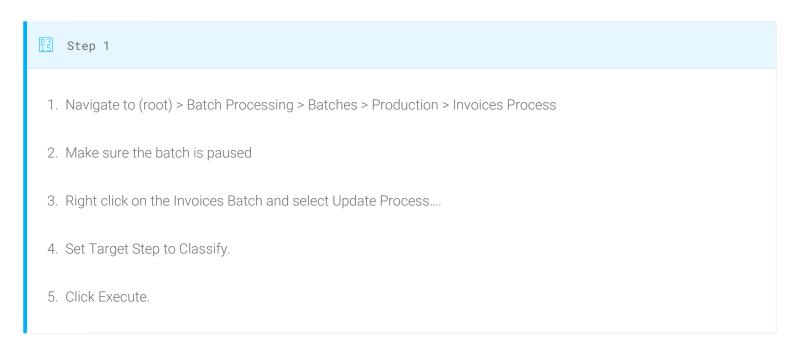


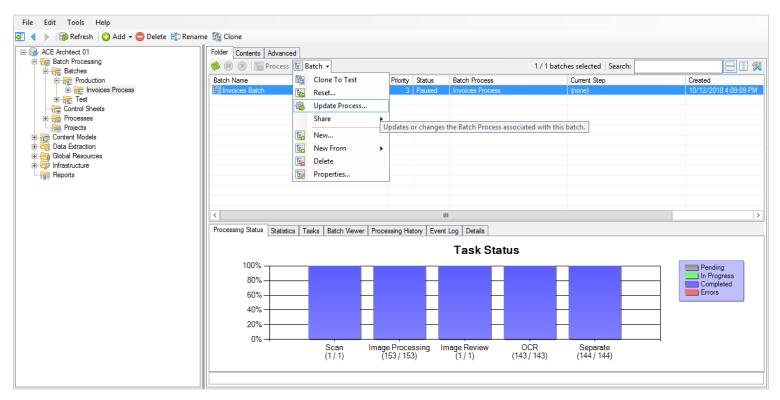
- 1. Add a new step
- 2. Set the Activity Type property of the new step to Classify Review
- 3. Set the Classification Viewer Settings > Content Model Scope to our Invoices Content Model
- 4. Save and Publish.

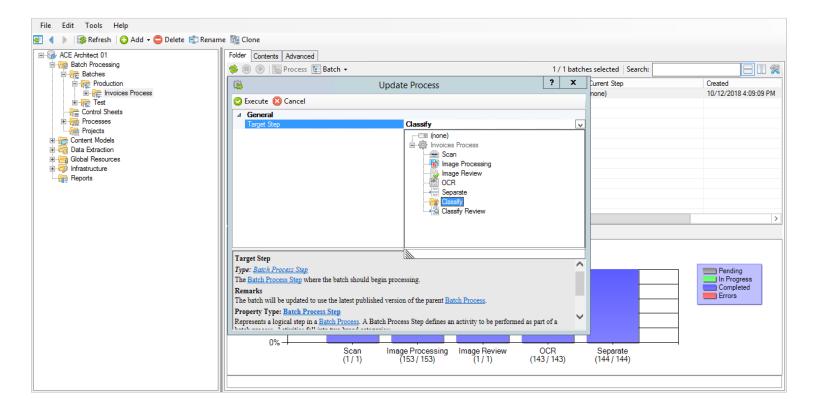


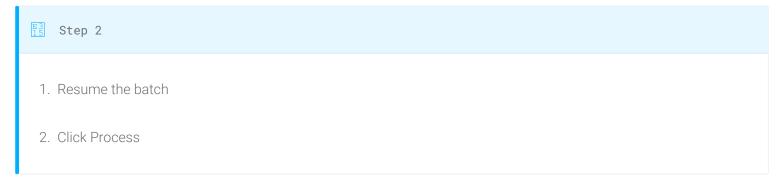


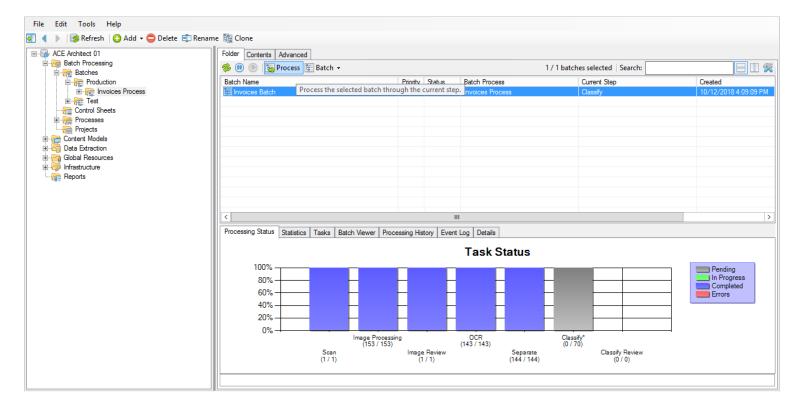
Updating the production batch











Reviewing and correcting classification



- 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

