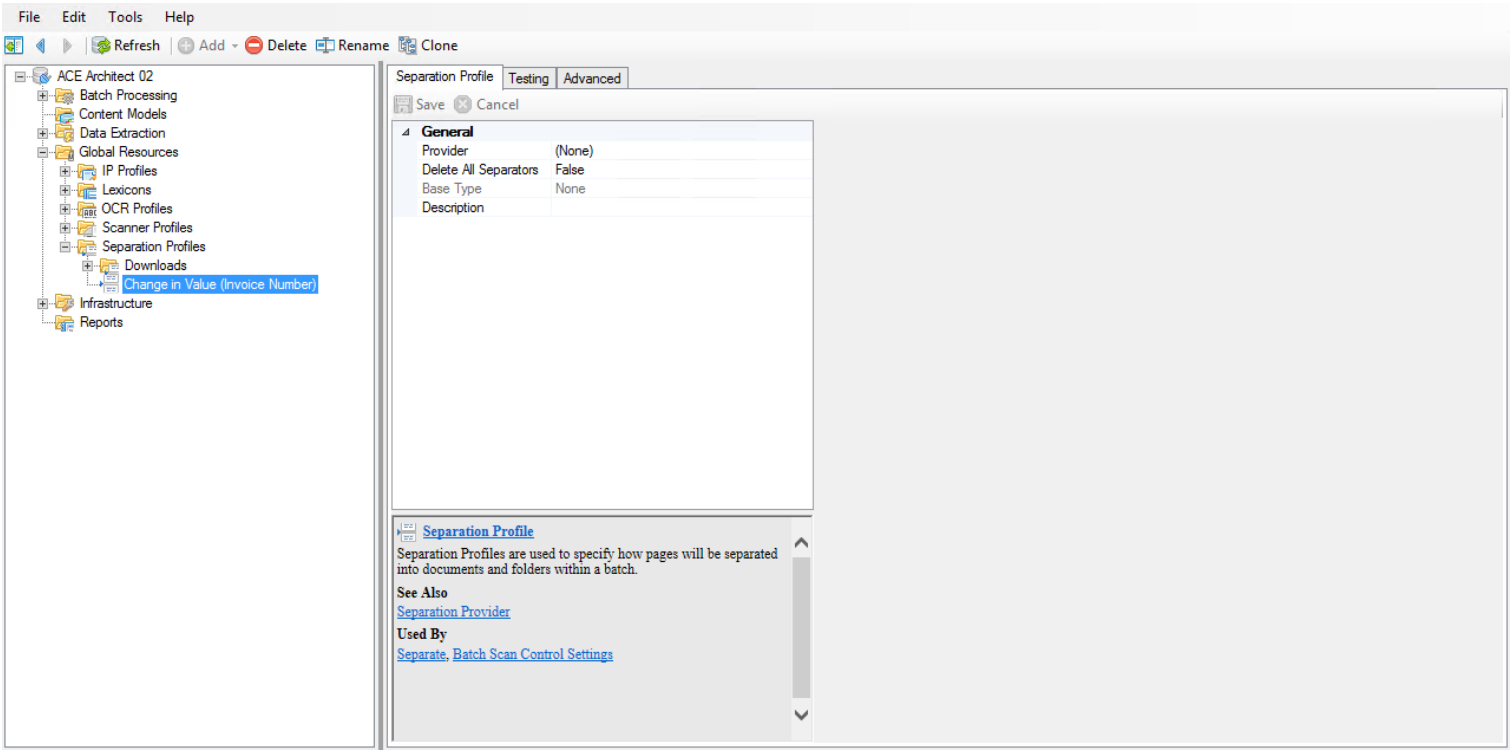
Organize

Separation

Creating a Separation Profile

> Global Resources / **Separation Profiles**

1. Right click and select `Add → Separation Profile...` .
2. Name it `Change in Value (Invoice Number)` .

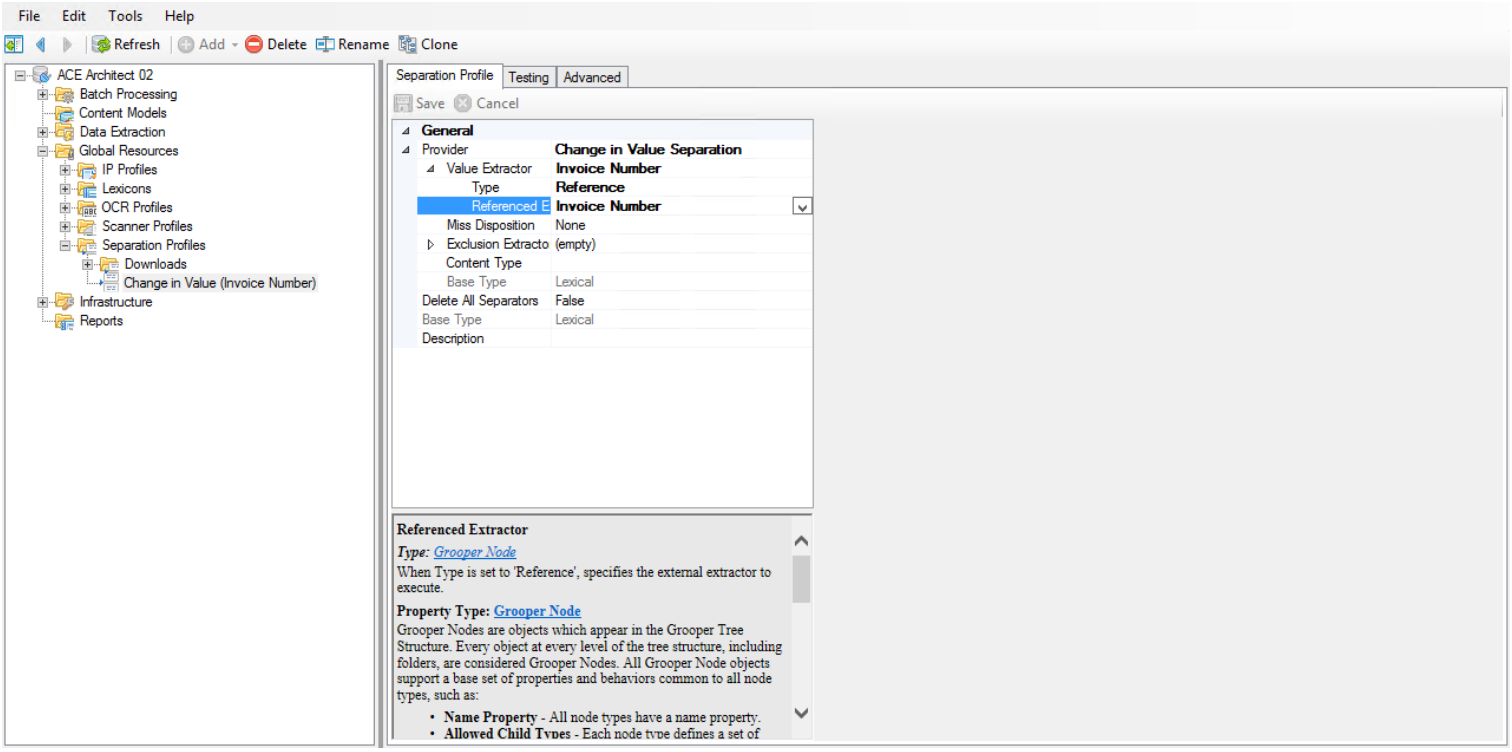


Configuring "Change in Value" separation

> Global Resources / Separation Profiles / **Change in Value (Invoice Number)**

Change the following property value(s):

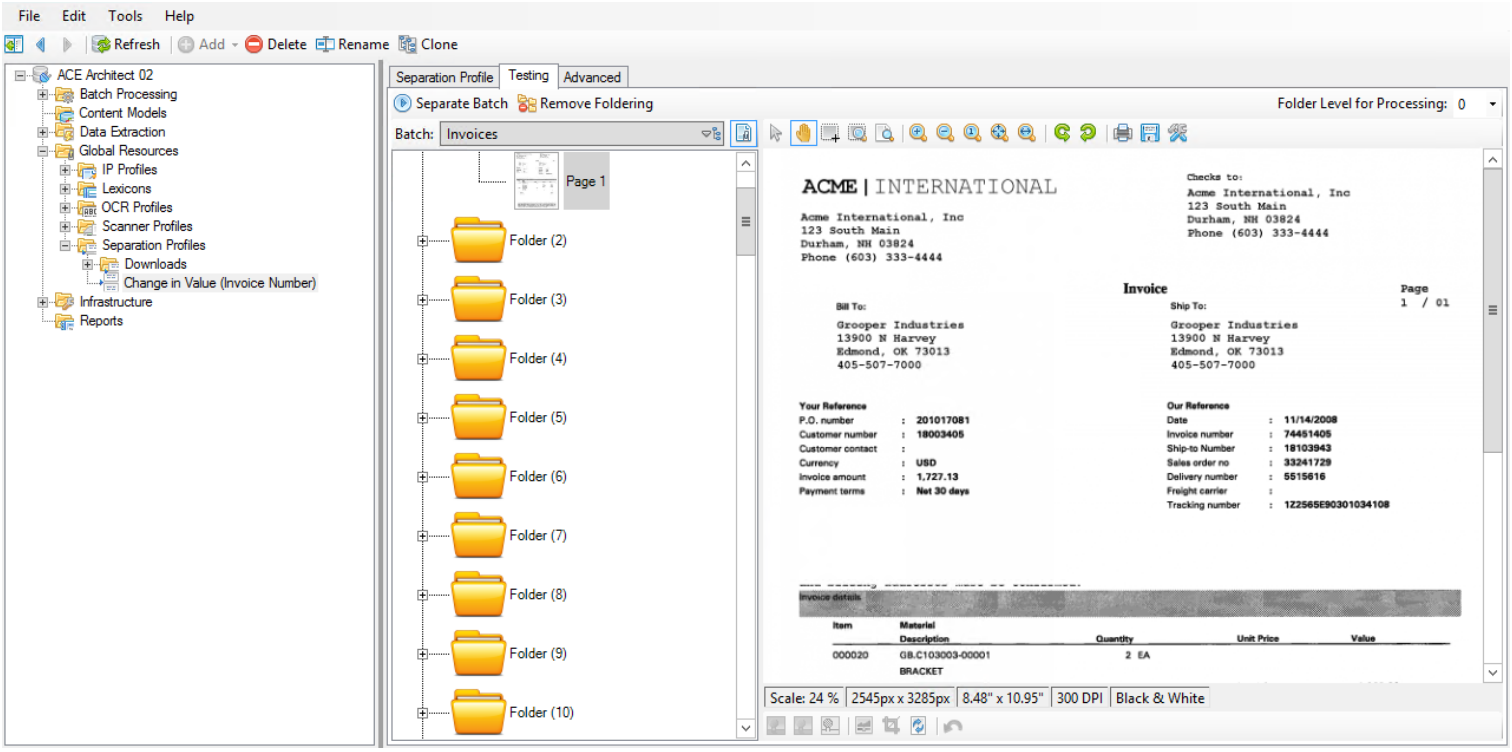
Property	Default Value	New Value
General		
Provider	-	Change in Value Separation
→ Value Extractor		
→ → Type	None	Reference
→ → Referenced Extractor	-	Invoice Number



Testing the Separation Profile

> Global Resources / Separation Profiles / **Change in Value (Invoice Number)**

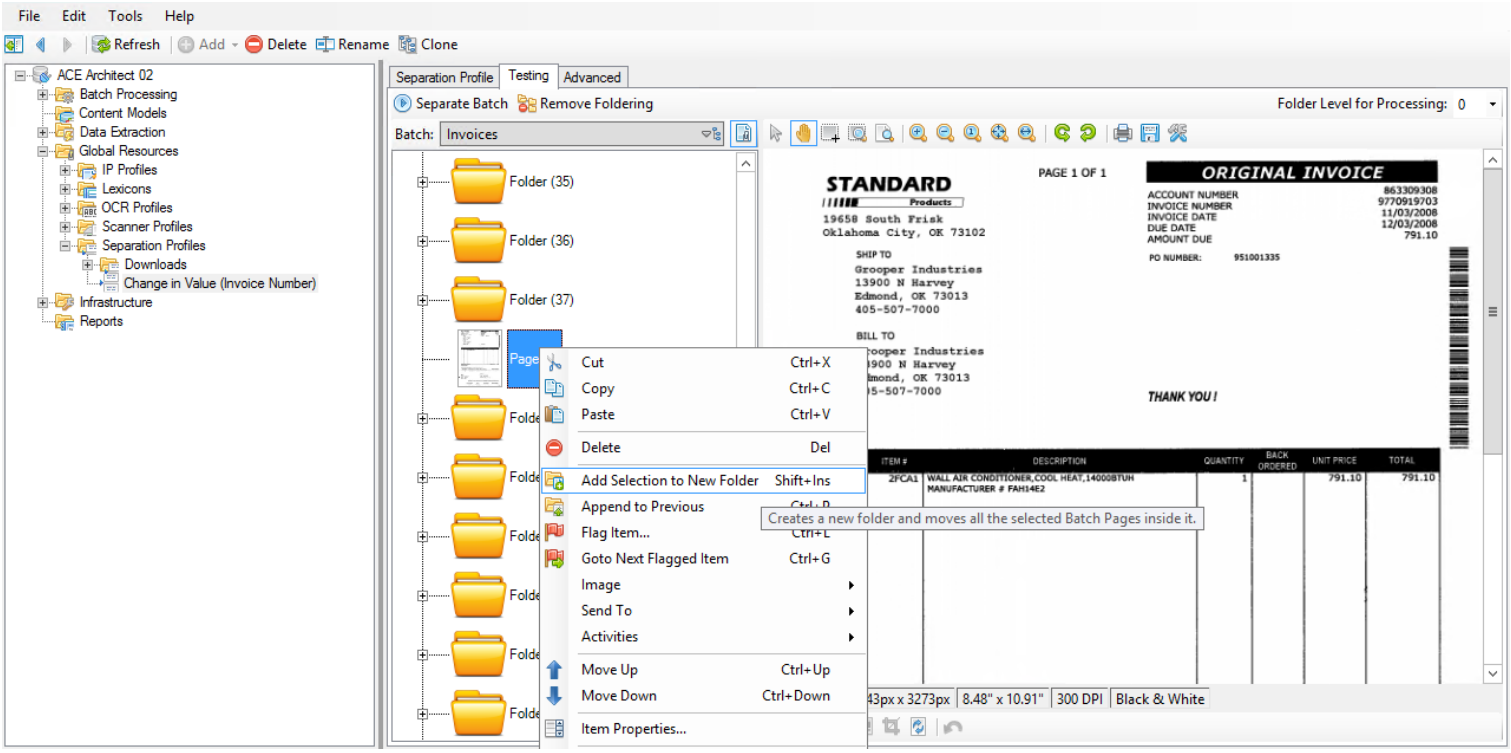
1. Switch to the `Testing` tab.
2. Press `Separate Batch`.



Correcting issues

> Global Resources / Separation Profiles / **Change in Value (Invoice Number)**

1. Select `Page 38`.
2. Right click and select `Add Selection to New Folder`.

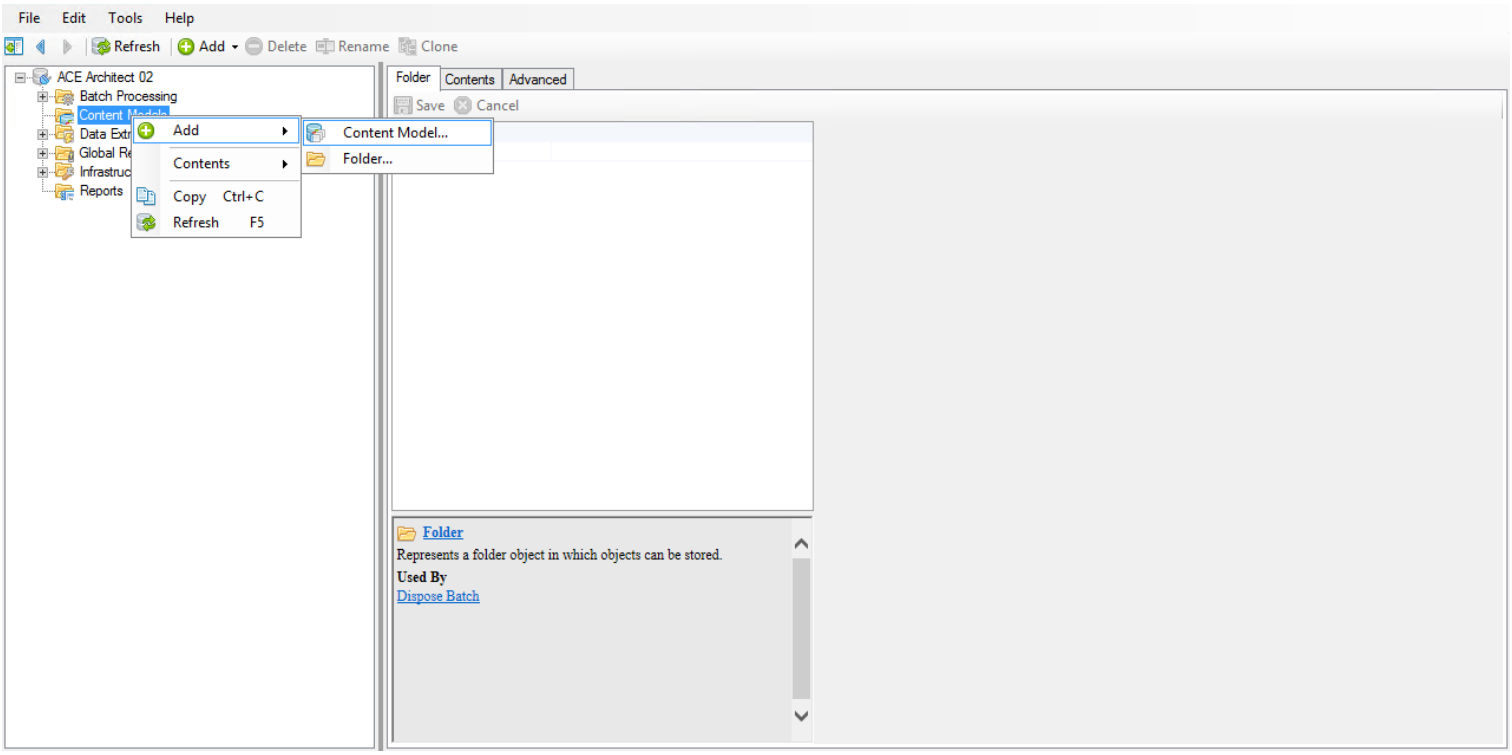


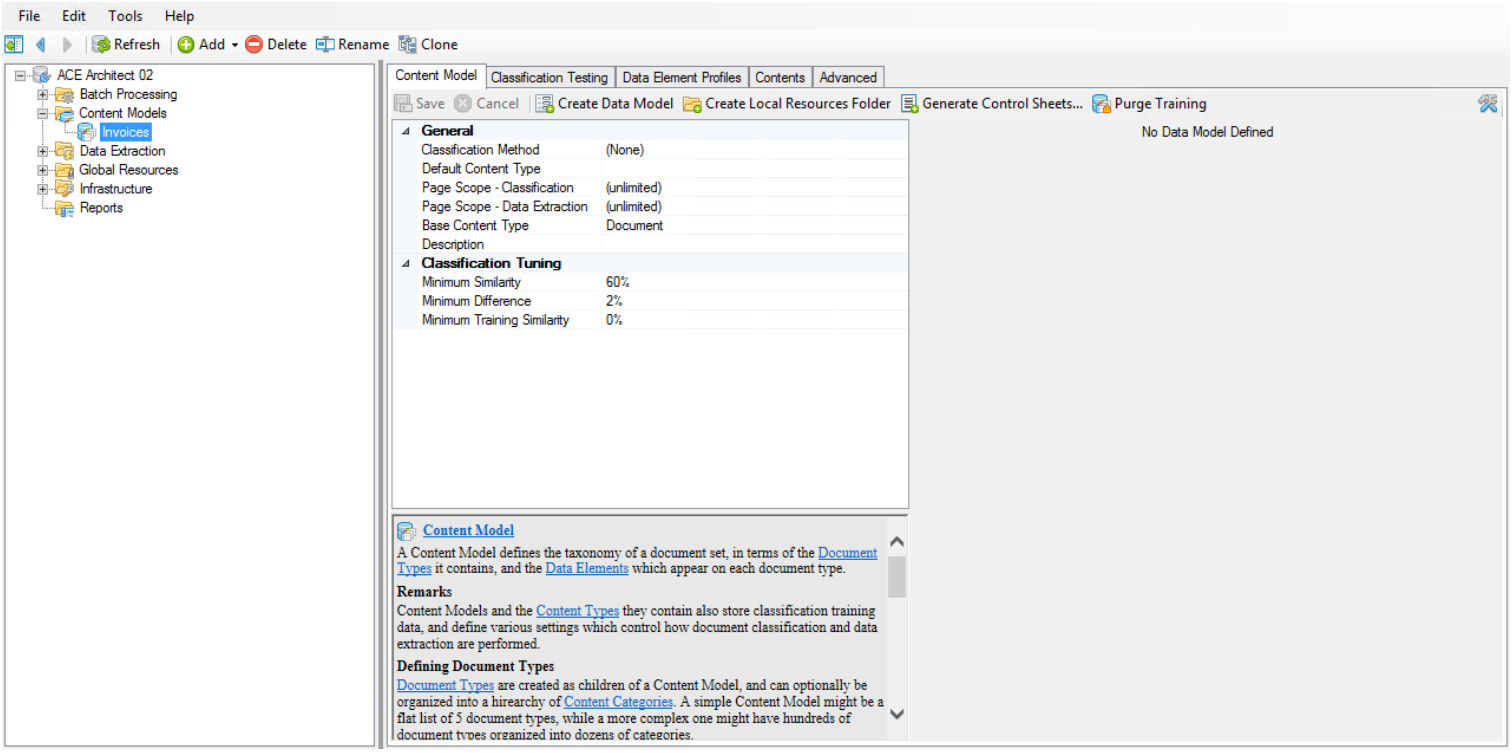
Classification

Setting up the Content Model

> **Content Models**

1. Right click and select `Add → Content Model...`.
2. Name it `Invoices`.

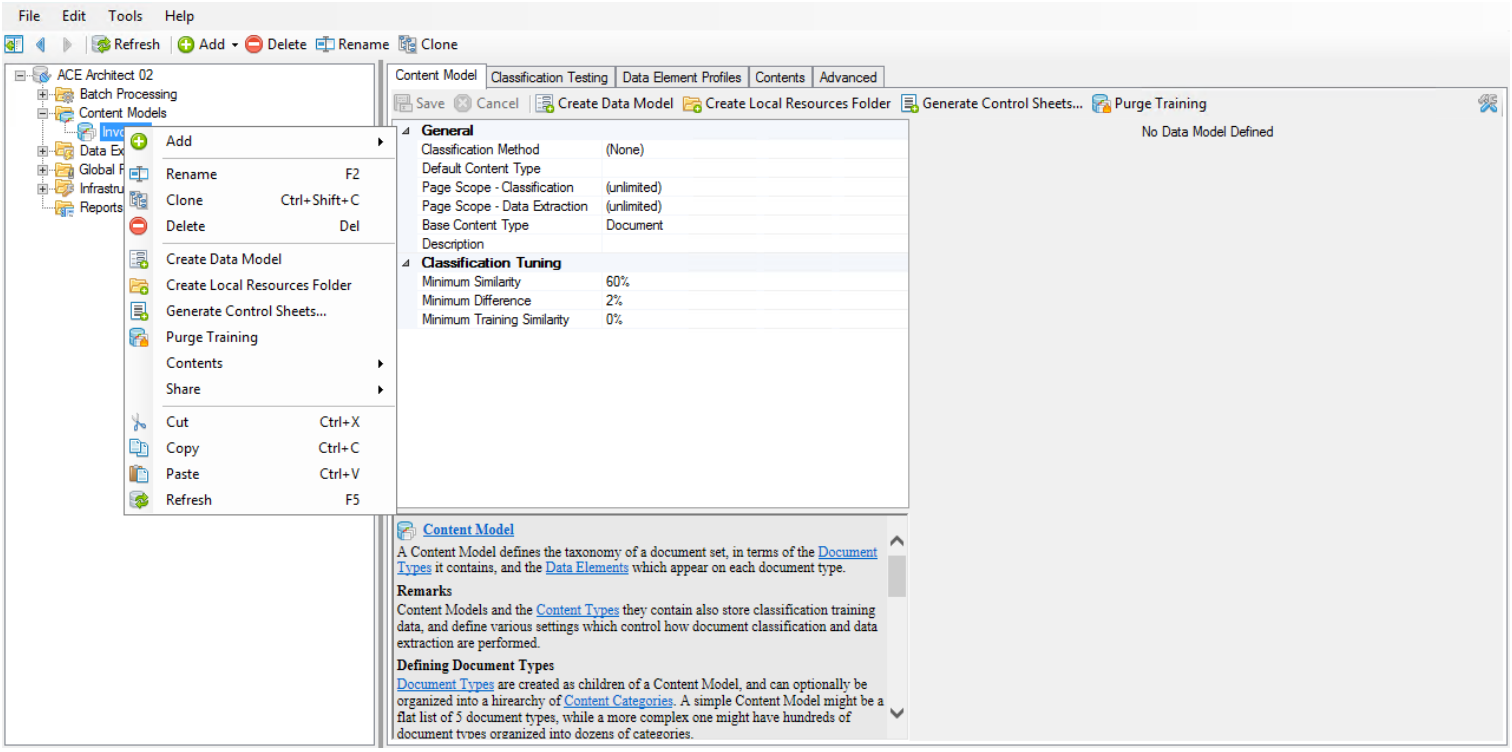

File
Edit
Tools
Help
Refresh
Add
Delete
Rename
Clone
ACE Architect 02
Batch Processing
Content Models
Data Extr
Global Re
Infrastruc
Reports
Add
Contents
Copy Ctrl+C
Refresh F5
Content Model...
Folder...
Folder
Contents
Advanced
Save
Cancel
Folder
Represents a folder object in which objects can be stored.
Used By
Dispose Batch

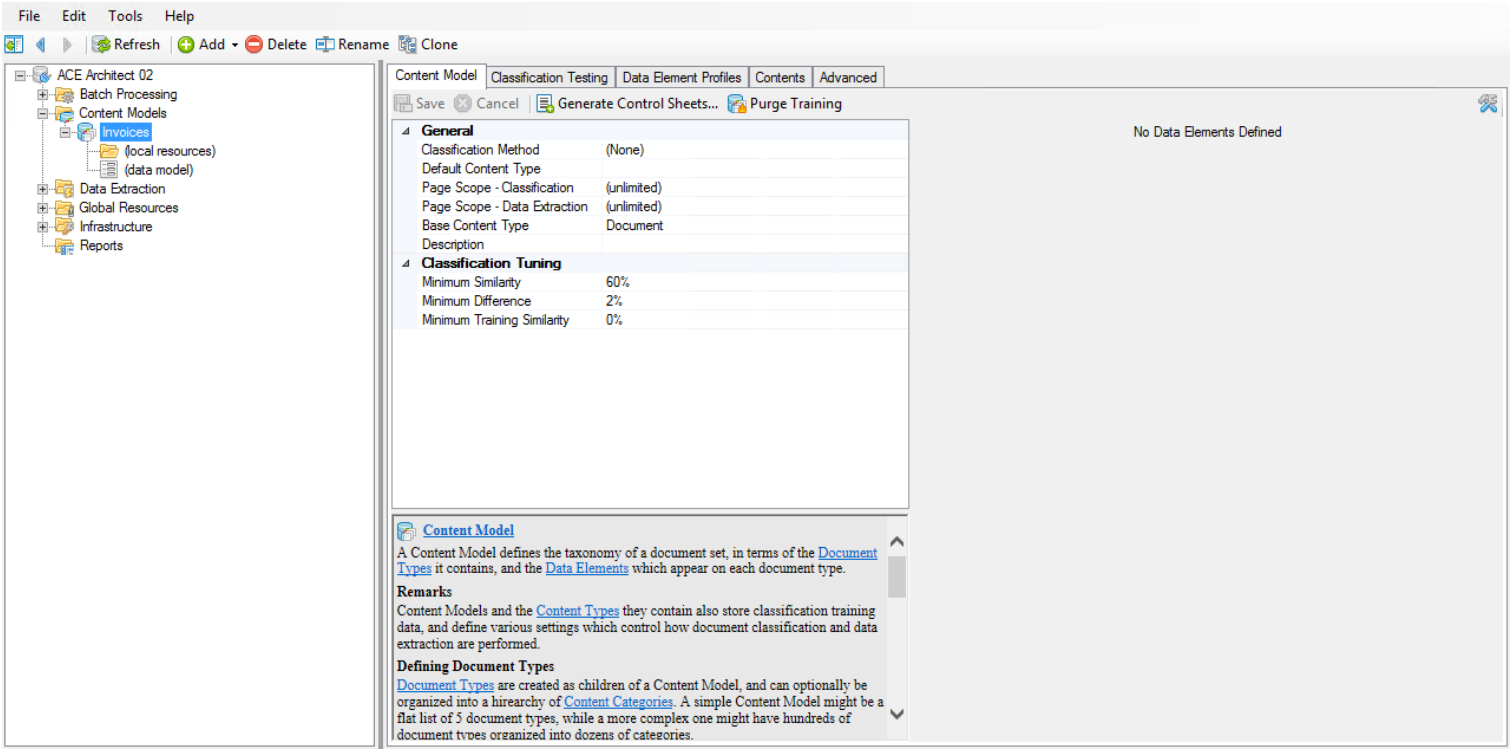

File
Edit
Tools
Help
Refresh
Add
Delete
Rename
Clone
ACE Architect 02
Batch Processing
Content Models
Invoices
Data Extraction
Global Resources
Infrastructure
Reports
Content Model
Classification Testing
Data Element Profiles
Contents
Advanced
Save
Cancel
Create Data Model
Create Local Resources Folder
Generate Control Sheets...
Purge Training
No Data Model Defined
General
Classification Method (None)
Default Content Type
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%
Content Model
A Content Model defines the taxonomy of a document set, in terms of the Document Types it contains, and the Data Elements which appear on each document type.
Remarks
Content Models and the Content Types they contain also store classification training data, and define various settings which control how document classification and data extraction are performed.
Defining Document Types
Document Types are created as children of a Content Model, and can optionally be organized into a hirearchy of Content Categories. A simple Content Model might be a flat list of 5 document types, while a more complex one might have hundreds of

Creating the Data Model and Local Resources folder

> Content Models / **Invoices**

1. Right click and select `Create Data Model`.
2. Right click and select `Create Local Resources Folder`.





Adding Document Types

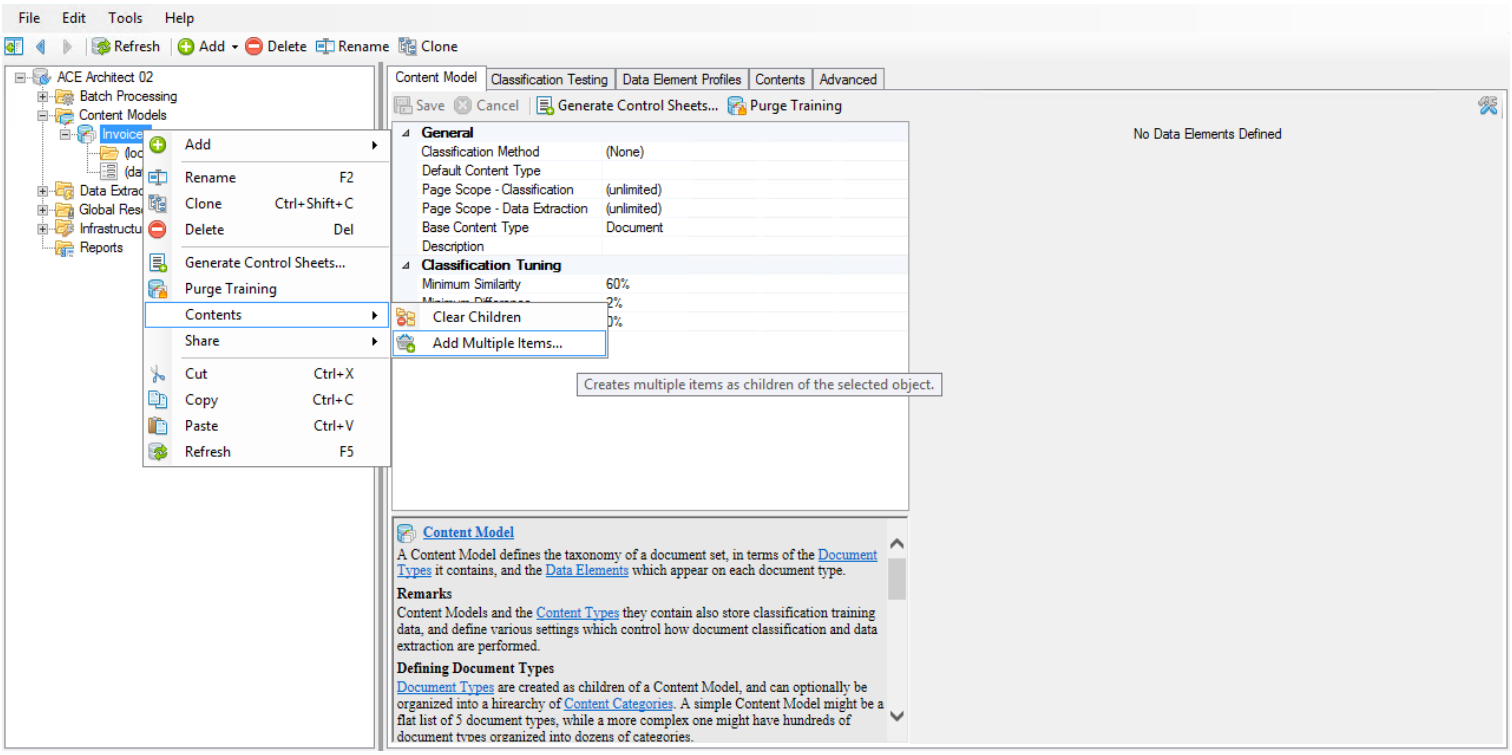
> Content Models / Invoices

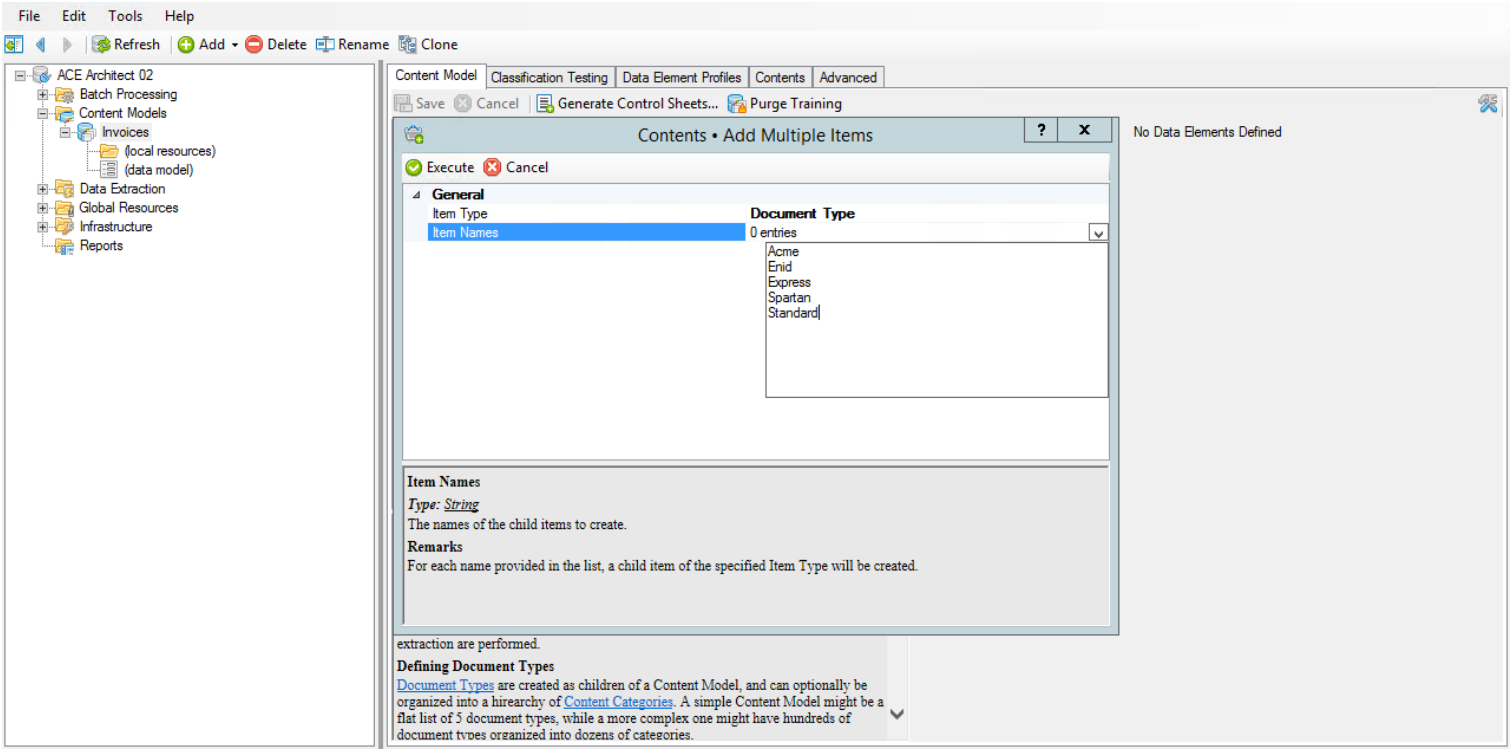
1. Right click and select `Contents → Add Multiple Items...`.
2. In the `Contents • Add Multiple Items` window, change the following property value(s):

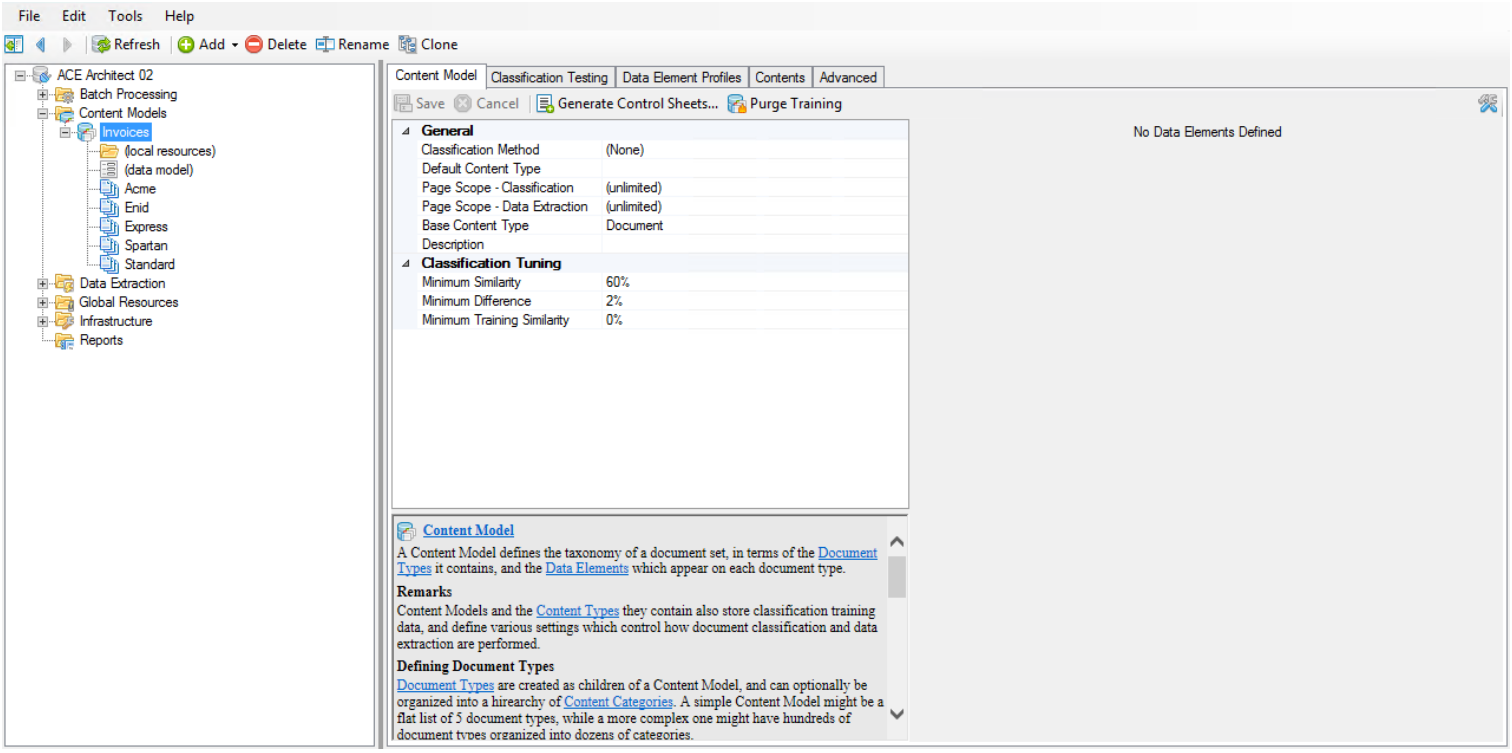
Property	Default Value	New Value
General		
Item Type	`Content Category`	`Document Type`

Item Names	-	`Acme`
		`Enid`
		`Express`
		`Spartan`
		`Standard`

3. Press `Execute`.


File Edit Tools Help
Refresh Add Delete Rename Clone
ACE Architect 02
Batch Processing
Content Models
Invoices
Data Extrac
Global Res
Infrastructu
Reports
Add
Rename F2
Clone Ctrl+Shift+C
Delete Del
Generate Control Sheets...
Purge Training
Contents
Share
Cut Ctrl+X
Copy Ctrl+C
Paste Ctrl+V
Refresh F5
Clear Children
Add Multiple Items...
Creates multiple items as children of the selected object.
Content Model
Classification Testing
Data Element Profiles
Contents
Advanced
Save Cancel Generate Control Sheets... Purge Training
General
Classification Method (None)
Default Content Type
Page Scope - Classification (unlimited)
Page Scope - Data Extraction (unlimited)
Base Content Type Document
Description
Classification Tuning
Minimum Similarity 60%
No Data Elements Defined
Content Model
A Content Model defines the taxonomy of a document set, in terms of the Document Types it contains, and the Data Elements which appear on each document type.
Remarks
Content Models and the Content Types they contain also store classification training data, and define various settings which control how document classification and data extraction are performed.
Defining Document Types
Document Types are created as children of a Content Model, and can optionally be organized into a hierarchy of Content Categories. A simple Content Model might be a flat list of 5 document types, while a more complex one might have hundreds of


File Edit Tools Help
Refresh Add Delete Rename Clone
ACE Architect 02
Batch Processing
Content Models
Invoices
(local resources)
(data model)
Data Extraction
Global Resources
Infrastructure
Reports
Content Model
Classification Testing
Data Element Profiles
Contents
Advanced
Save Cancel Generate Control Sheets... Purge Training
Contents • Add Multiple Items
Execute Cancel
General
Item Type Document Type
Item Names 0 entries
Acme
Enid
Express
Spartan
Standard
No Data Elements Defined
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.
extraction are performed.
Defining Document Types
Document Types are created as children of a Content Model, and can optionally be organized into a hierarchy of Content Categories. A simple Content Model might be a flat list of 5 document types, while a more complex one might have hundreds of

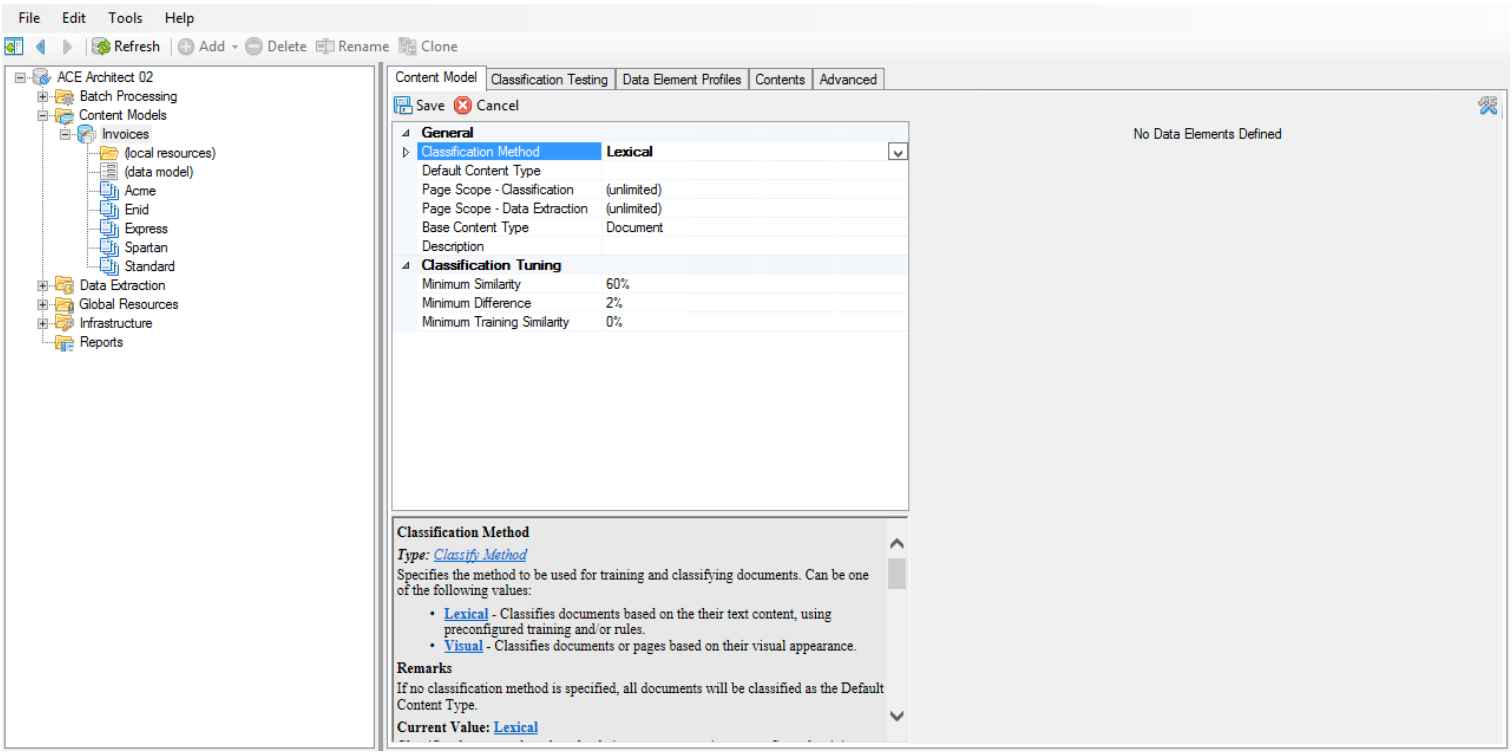


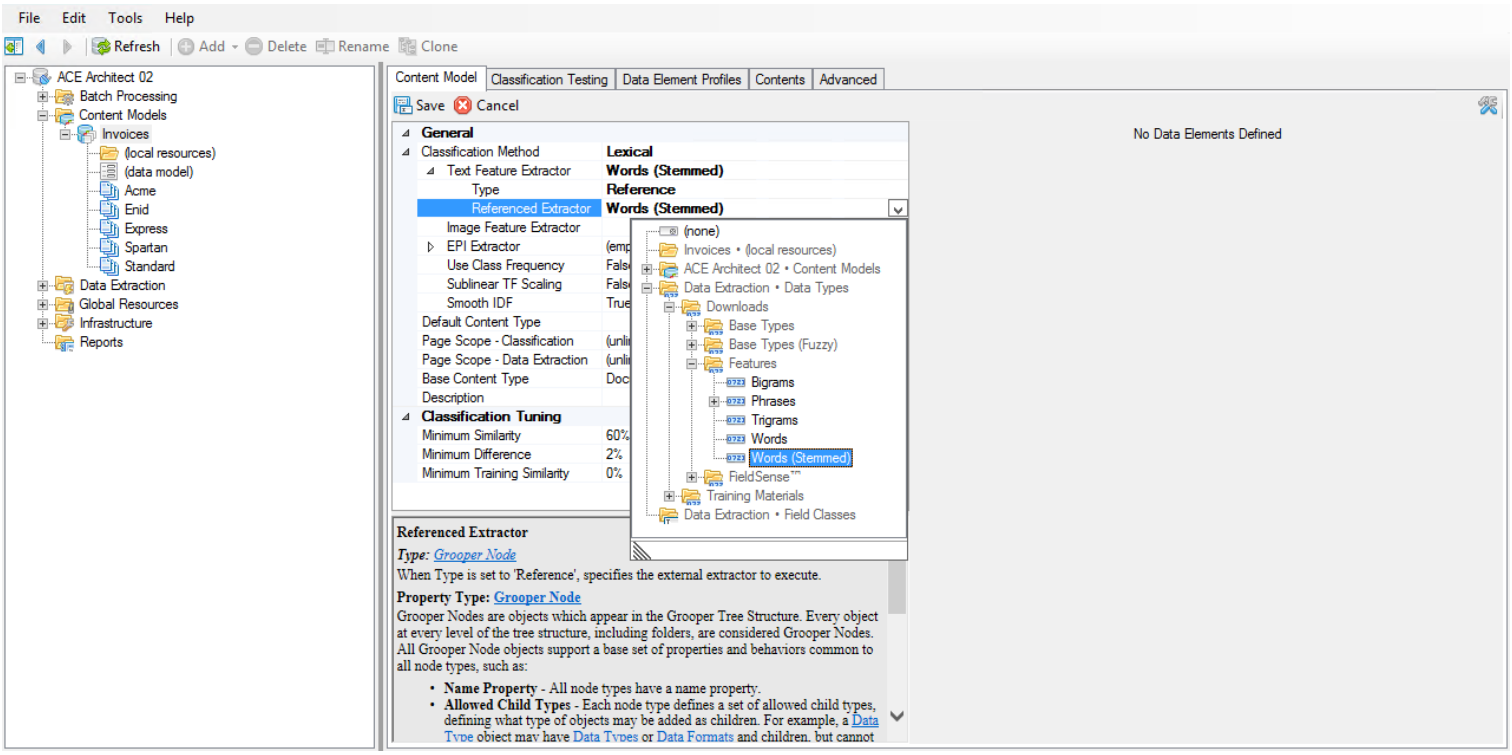
Setting the Classification Method

> Content Models / **Invoices**

Change the following property value(s):

Property	Default Value	New Value
General		
Classification Method	(None)	Lexical
→ Text Feature Extractor		
→ → Type	None	Reference
→ → Referenced Extractor	-	Words (Stemmed)


File Edit Tools Help
Refresh Add Delete Rename Clone
ACE Architect 02
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
General
Classification Method
Lexical
Default Content Type
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%
No Data Elements Defined
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 the 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

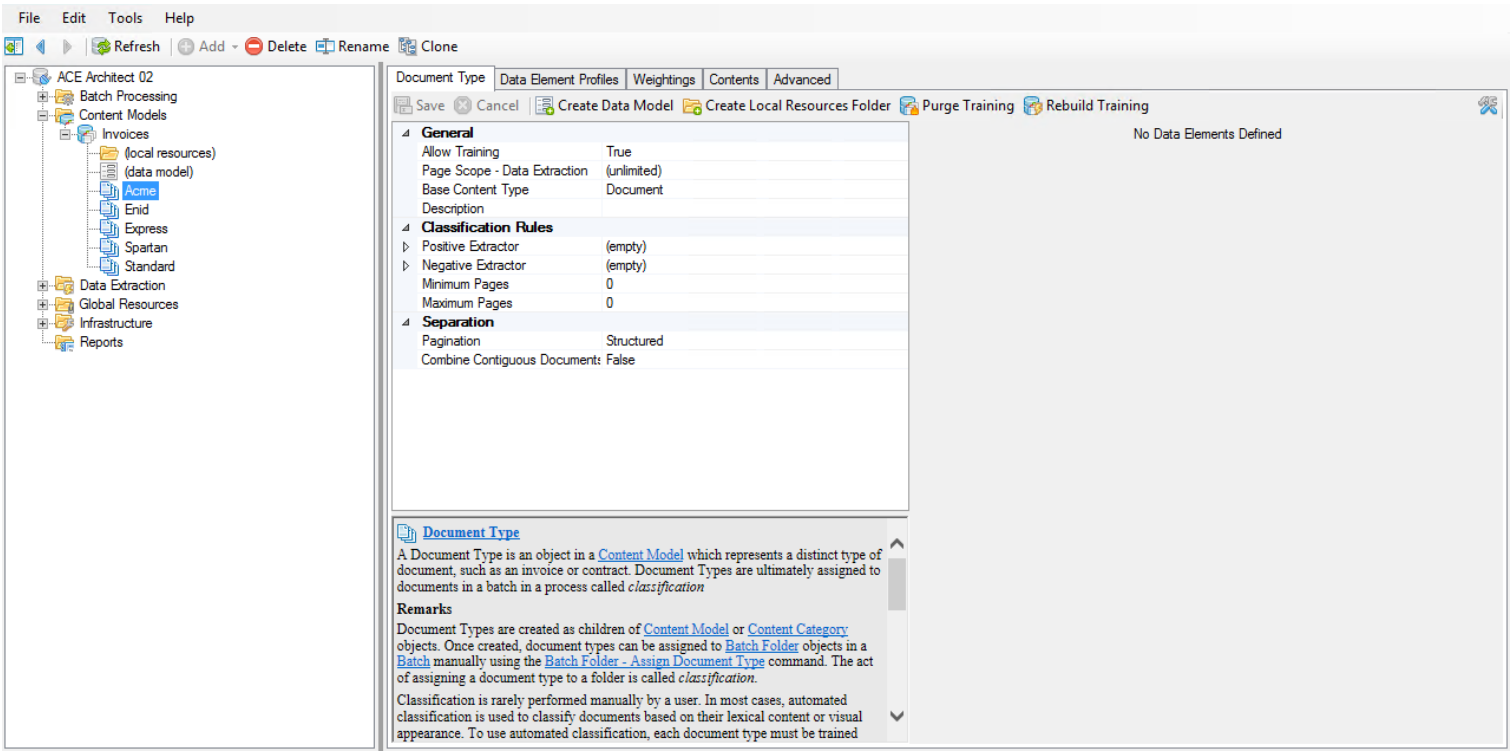

File Edit Tools Help
Refresh Add Delete Rename Clone
ACE Architect 02
Batch Processing
Content Models
Invoices
(local resources)
(data model)
Acme
Enid
Express
Spartan
Standard
Data Extraction
Global Resources
Infrastructure
Reports
Content Model
Classification Testing
Data Element Profiles
Contents
Advanced
Save
Cancel
General
Classification Method
Lexical
Text Feature Extractor
Words (Stemmed)
Type
Reference
Referenced Extractor
Words (Stemmed)
Image Feature Extractor
EPI Extractor
Use Class Frequency
Sublinear TF Scaling
Smooth IDF
Default Content Type
Page Scope - Classification
Page Scope - Data Extraction
Base Content Type
Description
Classification Tuning
Minimum Similarity
60%
Minimum Difference
2%
Minimum Training Similarity
0%
(none)
Invoices • (local resources)
ACE Architect 02 • Content Models
Data Extraction • Data Types
Downloads
Base Types
Base Types (Fuzzy)
Features
Bigrams
Phrases
Trigrams
Words
Words (Stemmed)
FieldSense™
Training Materials
Data Extraction • Field Classes
No Data Elements Defined
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

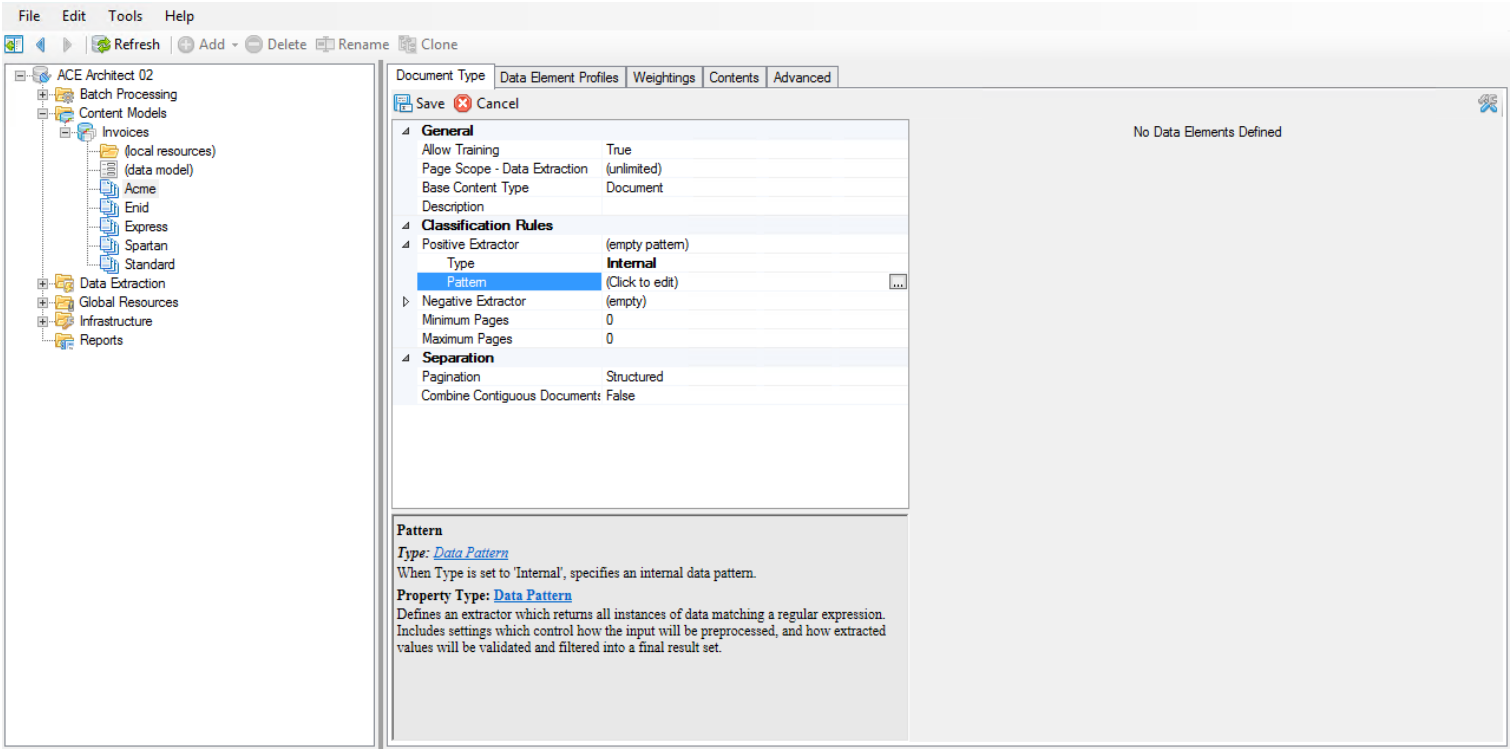
Setting a positive extractor for classification

> Content Models / Invoices / **Acme**

Change the following property value(s):

Property	Default Value	New Value
Classification Rules		
Positive Extractor		
→ Type	None	Internal
→ Pattern	*(Click to edit)*	*(Click to edit)*





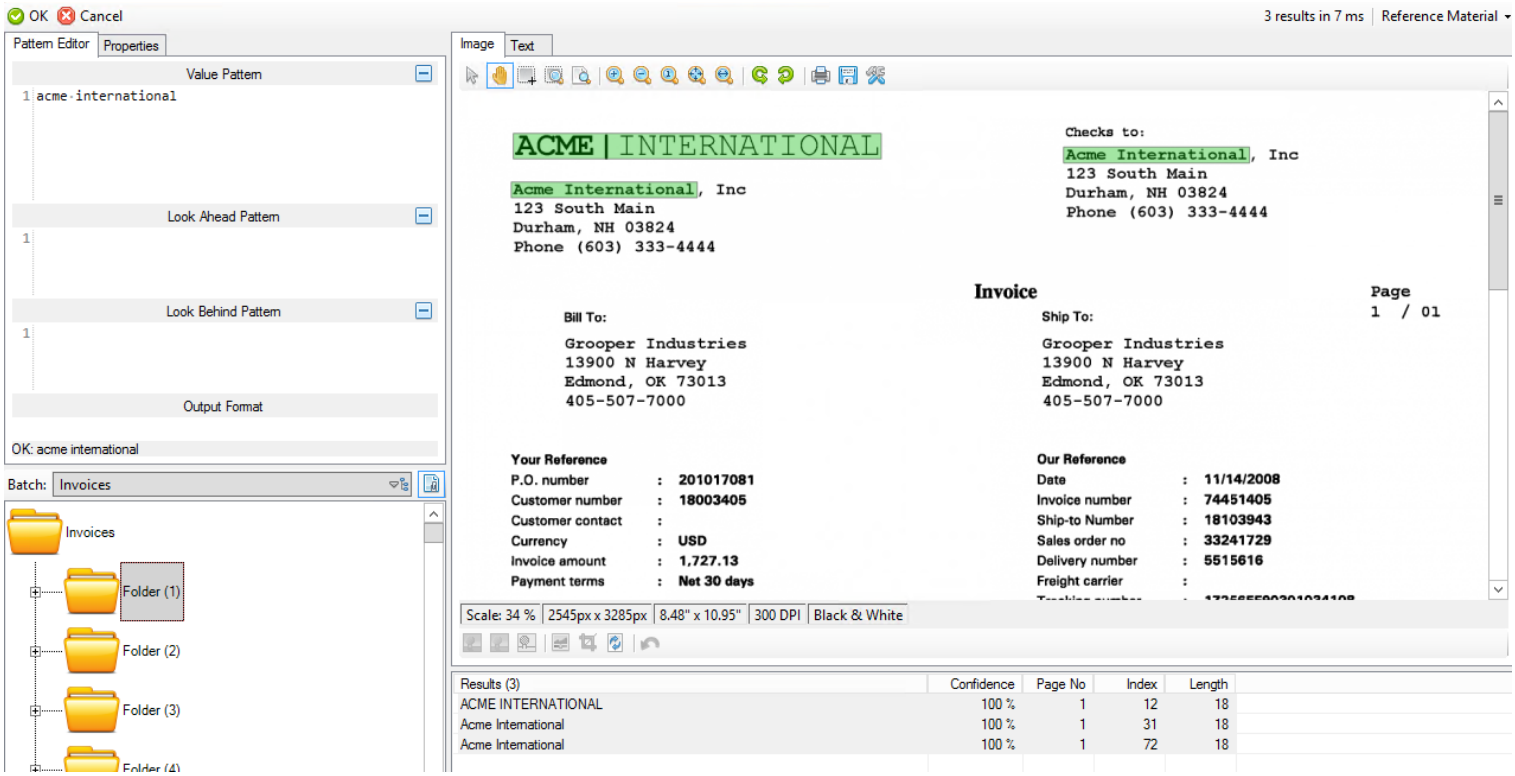
> Content Models / Invoices / **Acme**

1. Select `Folder (1)`.
2. Edit the pattern(s):

   Value Pattern

   ```
   acme international
   ```

3. Press `OK`.



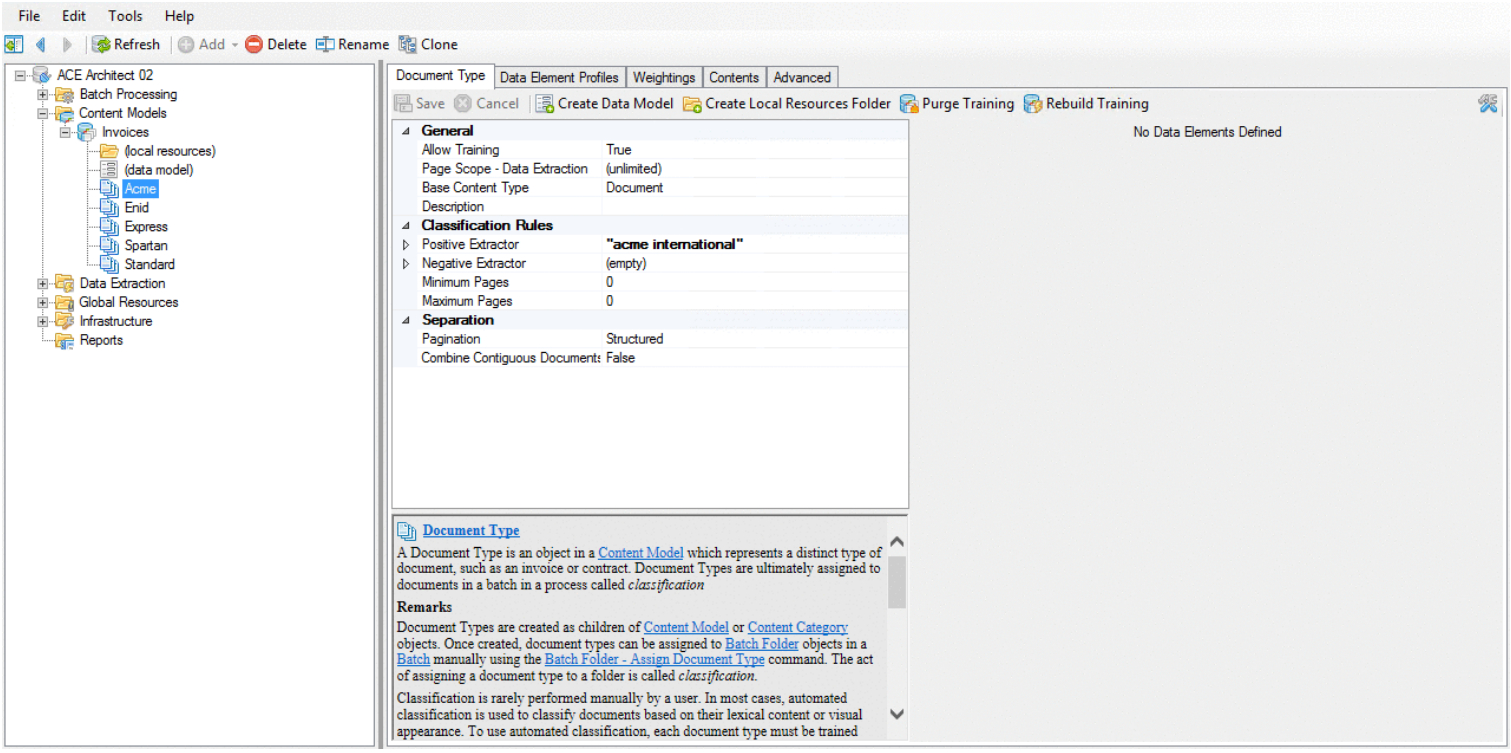
> Content Models / Invoices / *

Repeat the previous two steps for each of the other four vendors, using the following patterns in their respective `Value Pattern` editors:

Enid

```
enid parts
```

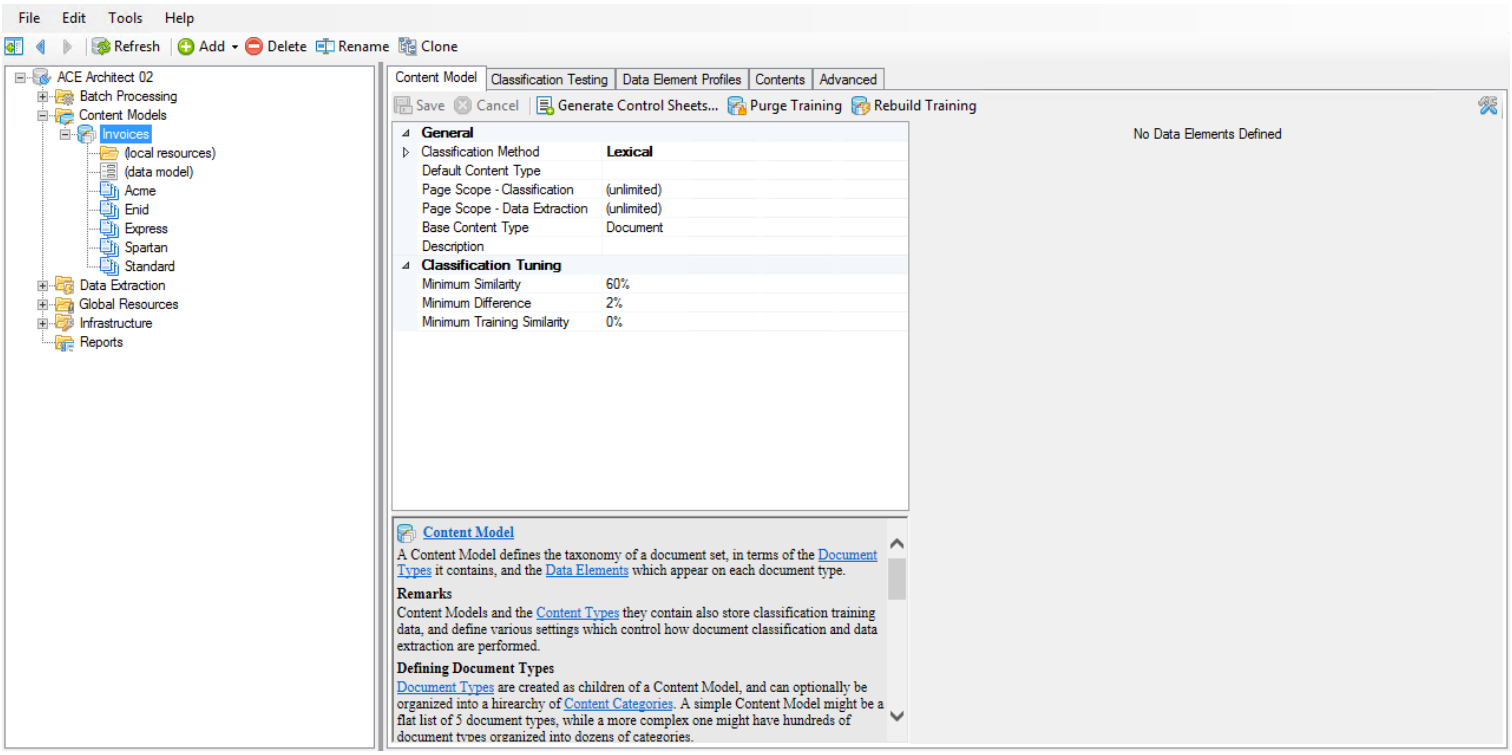
Express Spartan Standard

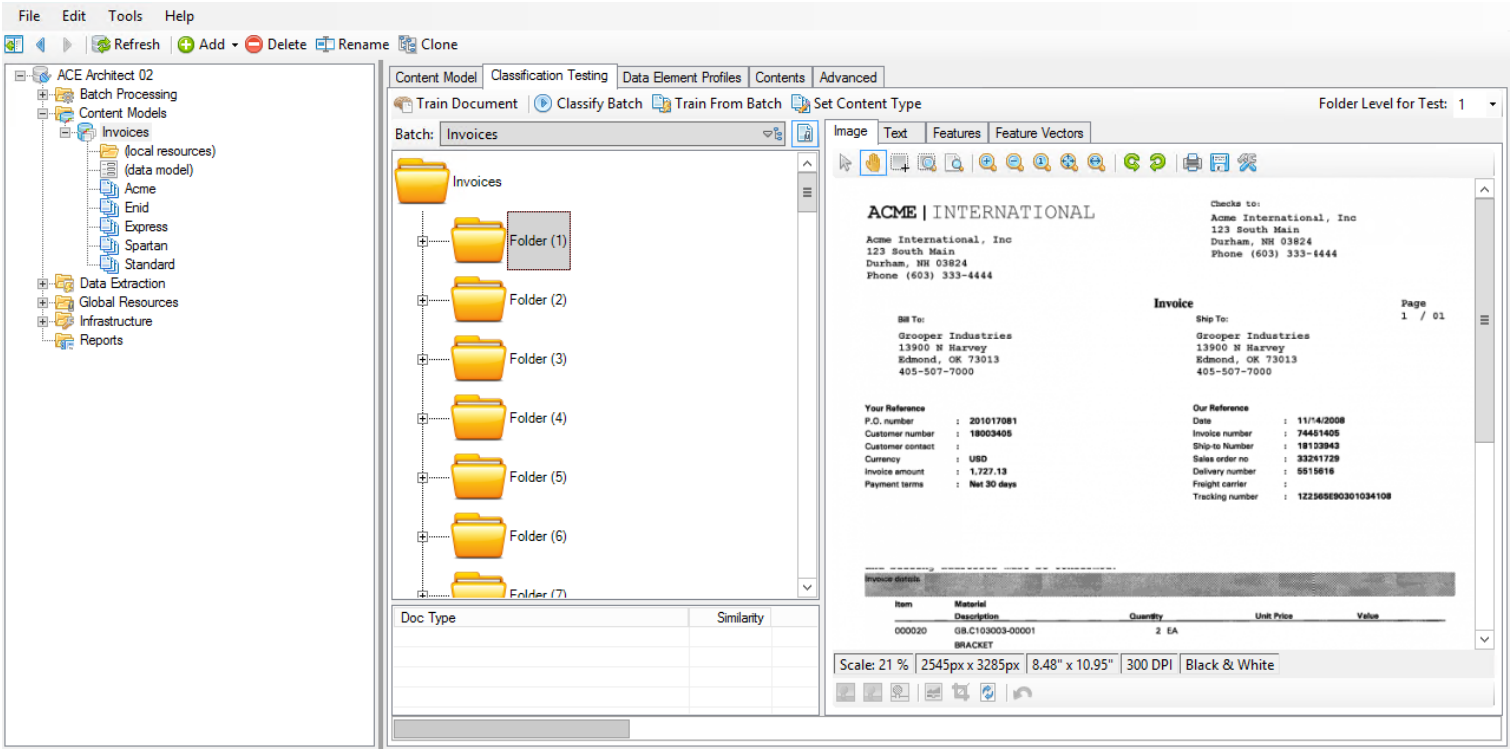


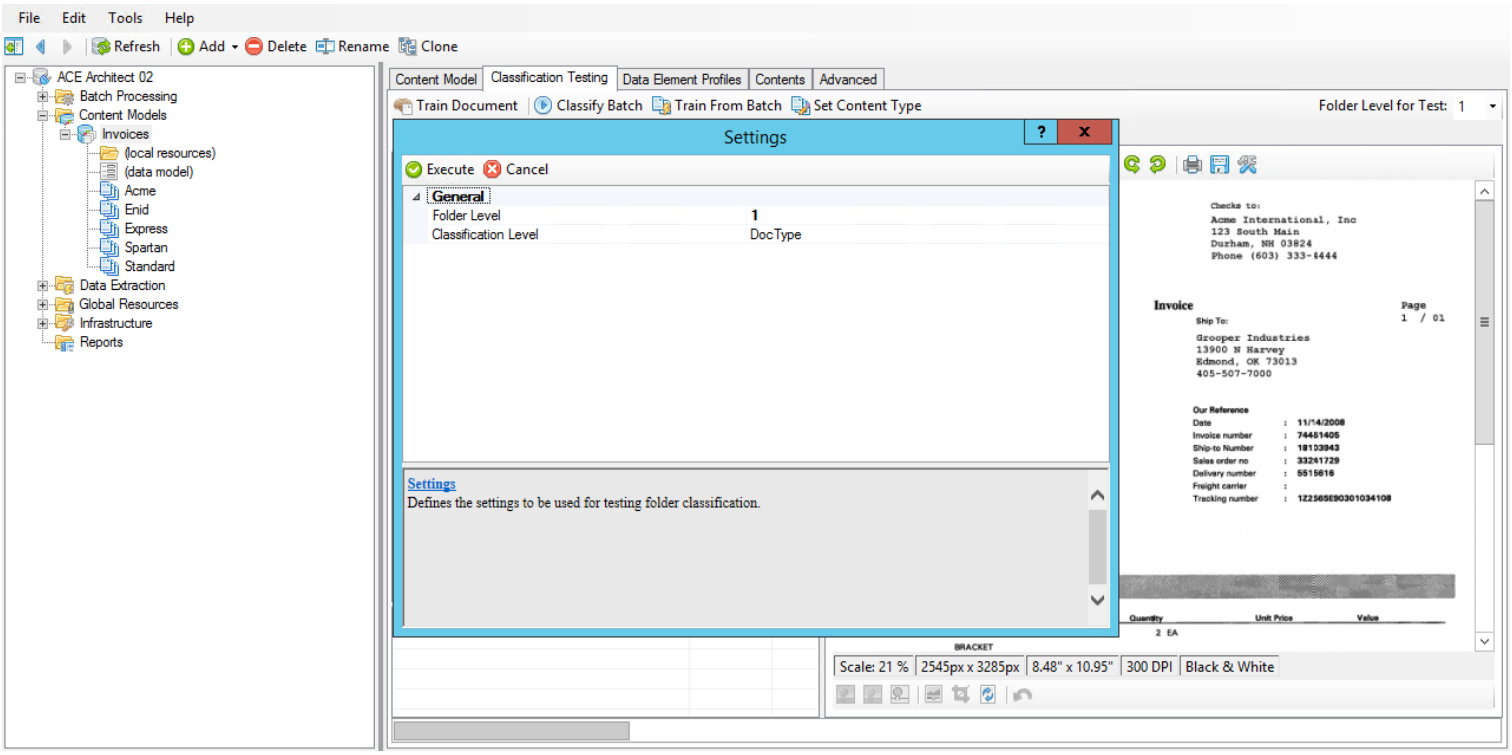
Testing classification

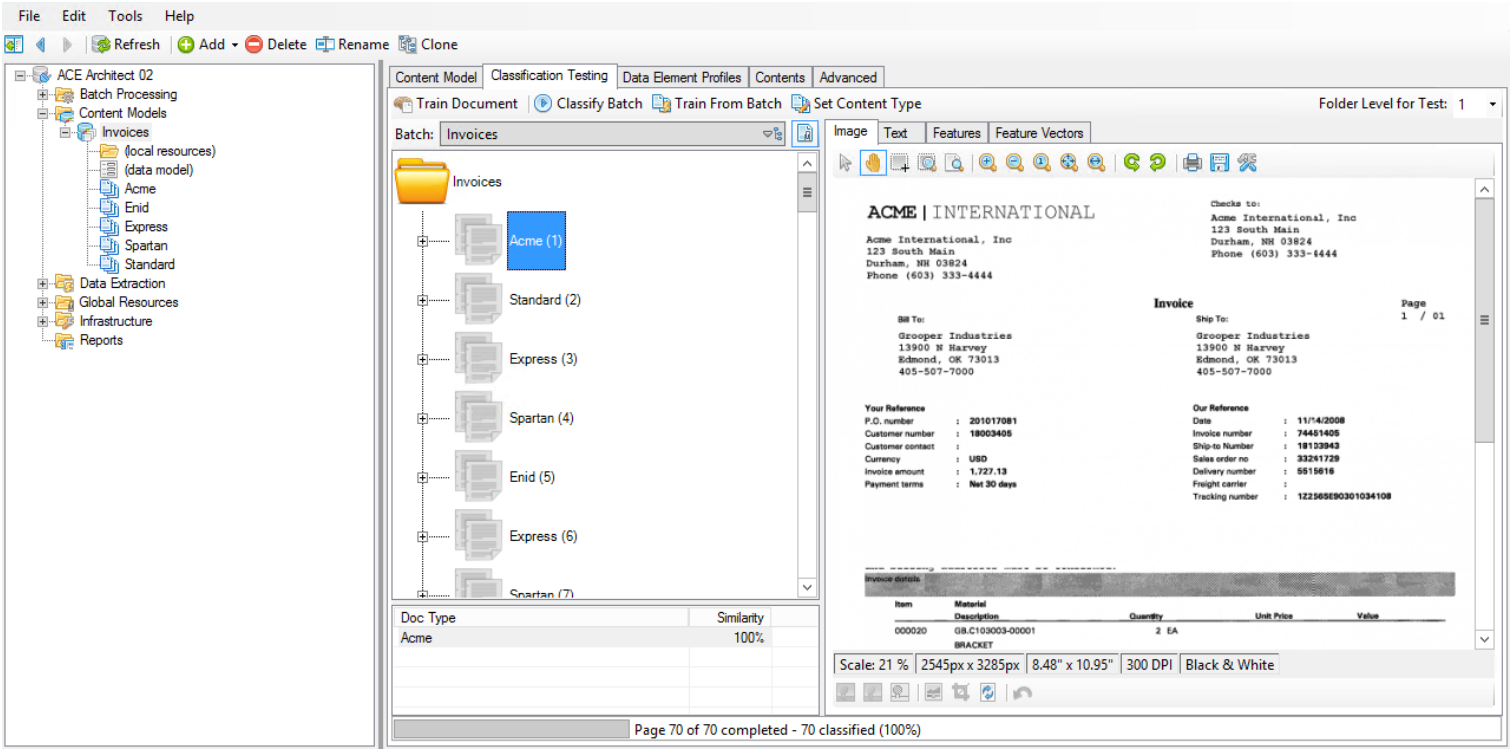
> Content Models / **Invoices**

1. Switch to the `Classification Testing` tab.
2. Press `Classify Batch`.
3. Press `Execute`.


File
Edit
Tools
Help
Refresh
Add
Delete
Rename
Clone
ACE Architect 02
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
General
Classification Method
Lexical
Default Content Type
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%
No Data Elements Defined
Content Model
A Content Model defines the taxonomy of a document set, in terms of the Document Types it contains, and the Data Elements which appear on each document type.
Remarks
Content Models and the Content Types they contain also store classification training data, and define various settings which control how document classification and data extraction are performed.
Defining Document Types
Document Types are created as children of a Content Model, and can optionally be organized into a hierarchy of Content Categories. A simple Content Model might be a flat list of 5 document types, while a more complex one might have hundreds of


File
Edit
Tools
Help
Refresh
Add
Delete
Rename
Clone
ACE Architect 02
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: Invoices
Invoices
Folder (1)
Folder (2)
Folder (3)
Folder (4)
Folder (5)
Folder (6)
Folder (7)
Doc Type
Similarity
Image
Text
Features
Feature Vectors
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) 333-6444
Invoice
Page
1 / 01
Bill To:
Grooper Industries
13900 N Harvey
Edmond, OK 73013
405-507-7000
Ship To:
Grooper Industries
13900 N Harvey
Edmond, OK 73013
405-507-7000
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 : 74451405
Ship-to Number : 18133943
Sales order no : 33241729
Delivery number : 5515616
Freight carrier :
Tracking number : 1Z2565E90301034108
Item
Material
Description
Quantity
Unit Price
Value
000020
GB.C103003-00001
BRACKET
2 EA
Scale: 21 %
2545px x 3285px
8.48" x 10.95"
300 DPI
Black & White


File Edit Tools Help
Refresh Add Delete Rename Clone
ACE Architect 02
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
Settings
Execute
Cancel
General
Folder Level 1
Classification Level DocType
Settings
Defines the settings to be used for testing folder classification.
Invoice
Scale: 21 %
2545px x 3285px
8.48" x 10.95"
300 DPI
Black & White


File Edit Tools Help
Refresh Add Delete Rename Clone
ACE Architect 02
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: Invoices
Invoices
Acme (1)
Standard (2)
Express (3)
Spartan (4)
Enid (5)
Express (6)
Doc Type
Similarity
Acme
100%
Image
Text
Features
Feature Vectors
ACME | INTERNATIONAL
Invoice
Scale: 21 %
2545px x 3285px
8.48" x 10.95"
300 DPI
Black & White
Page 70 of 70 completed - 70 classified (100%)