# KOFAX Transformation Modules 5

**Module 2 – Product Overview** 

What is it?



#### **Training Module Overview**

- •What is KTM?
- Structured, semi-structured and unstructured documents
- Classification, separation and extraction
- Kofax product differentiation
- Licensing
- Program Group components
- Integration with Kofax Capture

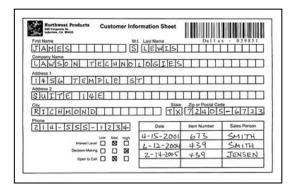


#### What is KTM?

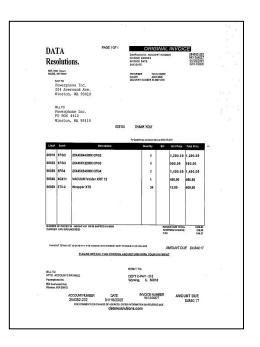
- KTM is a classification, separation and extraction "transformation" product for Kofax Capture (a suite of custom module add-ons)
- Although somewhat new to Kofax Capture, the product was introduced as Ascent Xtrata Pro 3.0. The product has now been rebranded as Kofax Transformation Modules, and this version is a 5.x release.
- Supports classification and extraction of structured, semi-structured and unstructured documents.
- Platform for delivering pre-configured, out-of-the-box solutions
- Offers learn-by-example extraction as well as numerous other methods to automatically extract data from documents
- Invoice classification by image and instructions (rules).

#### Structured or Unstructured?

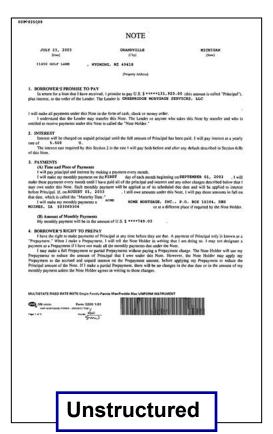
 KTM – Includes processing modules, a configuration application and several tools that work within Kofax Capture providing advanced document classification, separation, and extraction for structured, semi-structured and unstructured documents. It is optimized for invoice processing.



**Structured** 



Semi-structured





#### **Definitions**

- Classification Distinguishing between different Document Types and Form Types.
- Separation Creating boundaries between individual documents in a batch. Documents may be single-page or multi-page.
- Extraction The process of automatically lifting data from a document using rules or zones along with the output from a recognition engine (OCR, OMR, barcode, etc). This data usually is output to a database or document management application for later retrieval.

#### Some of KTM's Features Include...

- KCNS compatible
- Enterprise Option (High Availability) compatible
- Processes image (.tif) files
- 2 variations
  - Invoice Projects
  - Standard Projects
- Knowledge Bases
  - Generic and Specific Knowledge Bases
    - Country-specific
  - Based on "on-line learning"
  - Optional password protection
- "Professional" add-on available
  - Includes advanced page classification and trainable document separation using the special "Project Planner" configuration tool
- Other standard features
  - Database lookups
  - Regular expression support
  - Dictionary files
  - Table recognition
  - Scripting WinWrap Basic (similar to VBA)

#### KTM 4.0 – New Features

- Multiple instances of the KTM Server module
  - Classification and Separation
  - Extraction



- Document Review module
- Correction module: character-based correction
- Multiple instances of the KTM Validation module
- KTM Verification module
- Trainable Group Locator
- SQL Server support for statistics export
- Specific on-line learning for tables
- Page-level and zonal image cleanup
- Trainable Document Separation (Professional Version only)

#### KTM 4.5 – New Features

- Manual anchors for field registration
- 2. Web Validation and Verification modules
- 3. Arabic OCR engine (right □ left)
- 4. Upgraded ABBYY OCR engine
- 5. Integrated A2iA engine for cursive text and check recognition
- 6. Database (SQL and ODBC) lookup utility
- 7. Metadata-driven application setup
- 8. Vendor Locator
- 9. Invoice Evaluator
- 10. Country-specific settings and generic knowledge bases
- 11. Purchase order line item matching



#### **KTM 5.0 – New Features**

- Document Routing
- Batch Restructuring
- Multi Processor Support
- Thin client enhancements
  - New Thin Client Correction
  - Enhanced Thin Client Validation
    - folder support
    - batch editing
    - multiple steps
- Enhanced Document Review Module
- Script support in Verification
- Miscellaneous other features
  - New simplified licensing

KOFAX
Transformation
Modules 5





#### **Kofax Capture alone and with KTM Compared**

Feature	KC	KTM	Notes								
Functionality		10000000	100/100/24								
Classification - Image-based	Y	Y	Allows forms and non-content rich documents (e.g. checks, invoices, etc be classified. Kofax Capture based on histograms developed from single sample images. KTM uses multiple reduced-scal, low res sample pages allowing for variety within a form type.								
Classification - Rules-based	Y	Y	Allows content-rich documents (e.g. correspondence) to be classified. KC relies on form identification zones in fixed locations. KTM matches words anywhere on a page.								
Classification - Learn by Example	N	Y	Allows complex classification and is easy to extend.								
Separation - Rules-based	Y	Υ	Typically, only addresses simple separation schemes. KC relies manual separation, fixed pages, bar codes, patch coded and word(s) found in a fixe form ID zone. KTM classifies pages first and then separates based rules releated to page classification.								
Separation - Learn by Example	N	Y	The KTM Professional Add-on addresses complex, fully-automatic separation requirements based on trainable document separation (TDS) configurations.								
Extraction - template-based	Y	Y	Useful for structured documents and forms or semi-structured forms where required data is in the same location every time.								
Extraction - Free-form rules	N	Y	Required for semi-structured and unstructured documents.								
Extraction - Learn by example	N	Y	Trainable extraction using pre-created or user-created, transportable knowledge base files. Pre-configured methods for invoices. User-created methods for other kinds of documents.								
Automatic table extraction	Y (limited)	Υ	Kofax Capture only returns OCR data for the first row of a table. All other rows must be manually keyed. KTM performs automatic table data extraction of ALL rows and is trainable.								
Separate server instances for classification/separation and extraction	N	Y	KC offers a single Recognition Server module that may be used for simple classification, separation and extraction. KTM offers the option of two KTM Servers: one typically used for classification and separation and the second used for extraction.								
Multiple instances of Validation	N	Y	KTM offers up to five instances of KTM Validation as well as a separate KTM Verification module.								

#### Licensing

- New simplified licensing
  - New base licensing scheme based on numbers of fields extracted
    - 5-Fields Extraction Base License
    - Unlimited Field Extraction Base License
- Both base licenses include
  - Layout and content classification
  - Fixed- and free-form extraction
  - Invoice header and amount extraction (GLs/IHL)
  - Keying modules (inc web clients)

#### **Licensing – Add-ons**

#### New simplified licensing

- Add-ons
  - Table Extraction
  - Professional (Project Planner)
  - Additional OCR engines (Kadmos)
  - Check Recognition
  - Cursive Field Recognition
  - Cursive Full-Page Recognition
  - Arabic OCR
  - Invoice Add-On Pack (Vendor Locator & Invoice Evaluator)
  - Invoice Line Items Matching
  - Medical Claims Add-on Pack

#### **Licensing - Features**

#### New simplified licensing

Name	Base /	volume /	Classif	fication	Extraction							Add Ons								Usage	
	AddOn	system	Content	Layout	LIM Locator	Vendor Locator	Invoice Evaluator	Invoice-sp ecific Group Locators	Table Locator	IHL	all other Locators	Keying Modules	Kadmos	Check Recog.	Cursive Recog.	Document Reader	Arabic	Thin Clients	Doc Sep	Medical Claims project	KTT developm ent
KTM 5-Fields Extraction	Base	volume license	х	Х				х		х	х										
KTM Unlimited Fields Extraction	Base	volume license	х	Х				х		х	х										
KTM Table Extraction	add on	volume license							Х												
KTM Professional	add on	volume license																	Х		
KTM Keying Modules	add on	system license										Х									
KTM Additional OCR Engines	add on	volume license											х								
KTM Check Recognition	add on	volume license												Х							
KTM Cursive Field Recognition	add on	volume license													х						
KTM Document Reader	add on	volume license														х					
KTM Arabic OCR	add on	volume license															х				
KTM Invoice Add-on Pack	add on	volume license				х	х														
KTM Invoice Line Item Matching	add on	volume license			х																
KTM Web Server	add on	system license																х			
KTM Medical Claims Add-on Pack	add on	system license																		Х	

#### **Licensing – New "Mix & Match"**

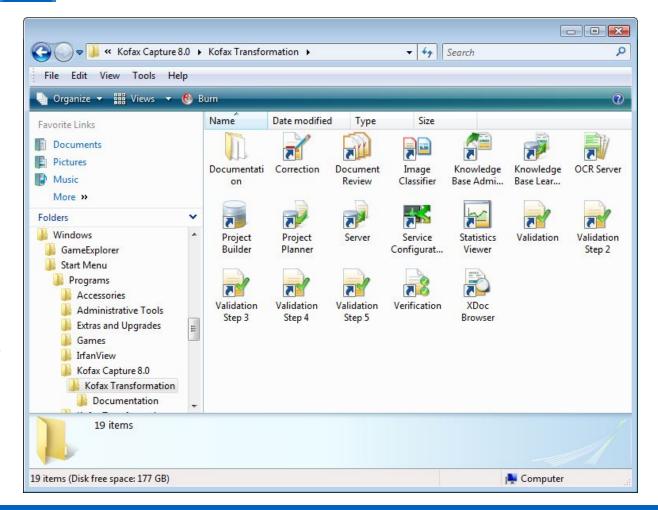
- New simplified licensing
- 'Mix-and-match' possible
  - Mixture of base license volumes for a single system
     900k total volume = 300k 5-field extraction + 600k unlimited field extraction
  - Different volumes for add-ons possible
     30k Table Extraction + 120k Arabic OCR



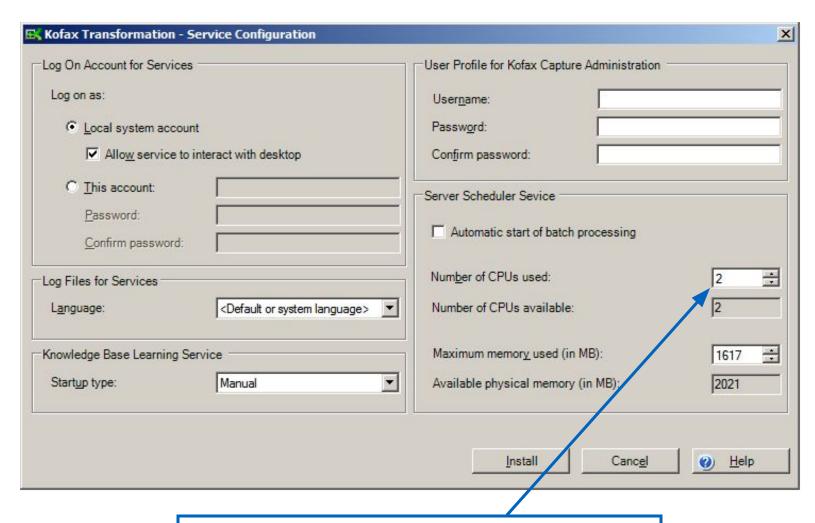
#### **Program Group Components**

### KOFAX Transformation Modules 5

NOTE: In this
example, 5
instances of KTM
Validation have
been installed.
Project Planner
(Professional
Upgrade) has also
been installed.



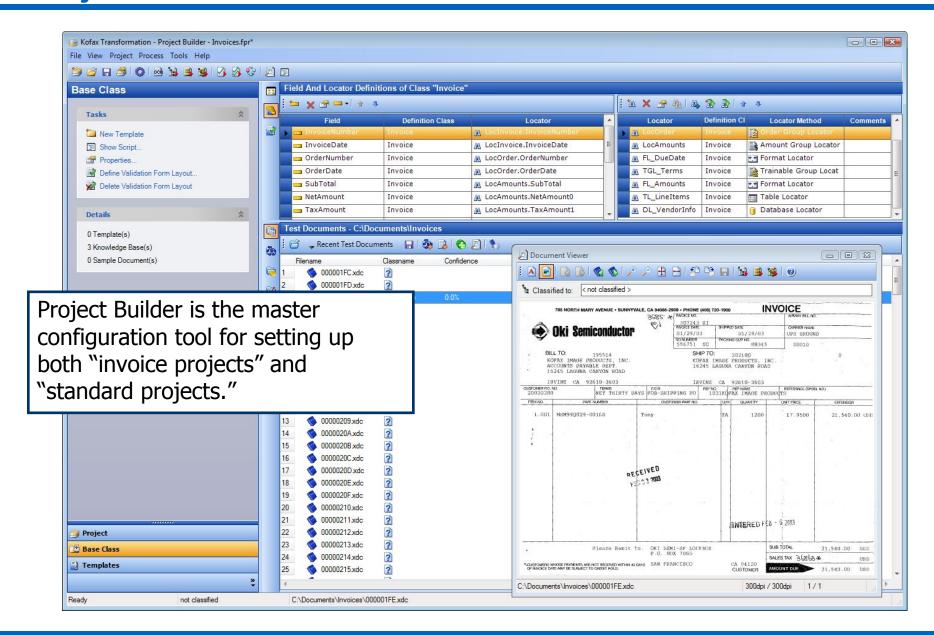
#### **KTM Service Configuration**



KTM 5 will support up to 16 CPU cores!



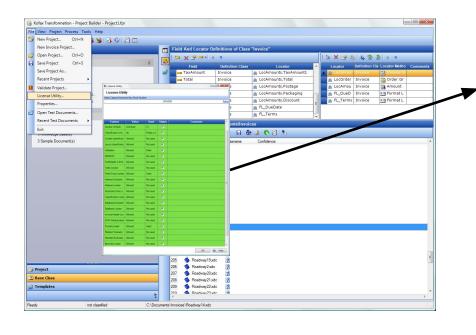
#### **Project Builder**



#### **KTM License Utility**

KTM has its own license utility. It is available from the "File" menu in Project Builder, and allows you to view your license values and what you are using for any project.

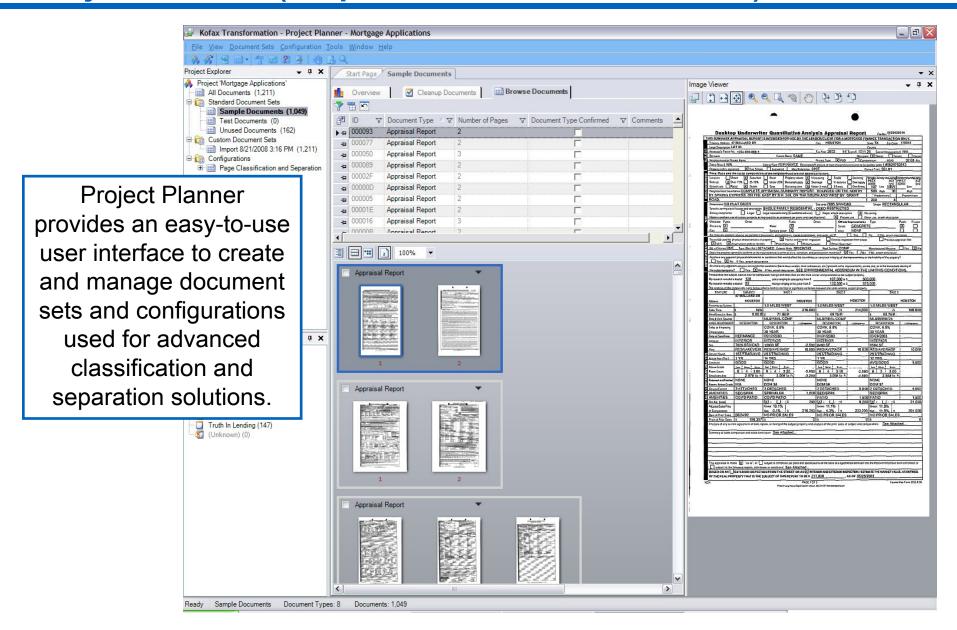
More about Project Builder follows...





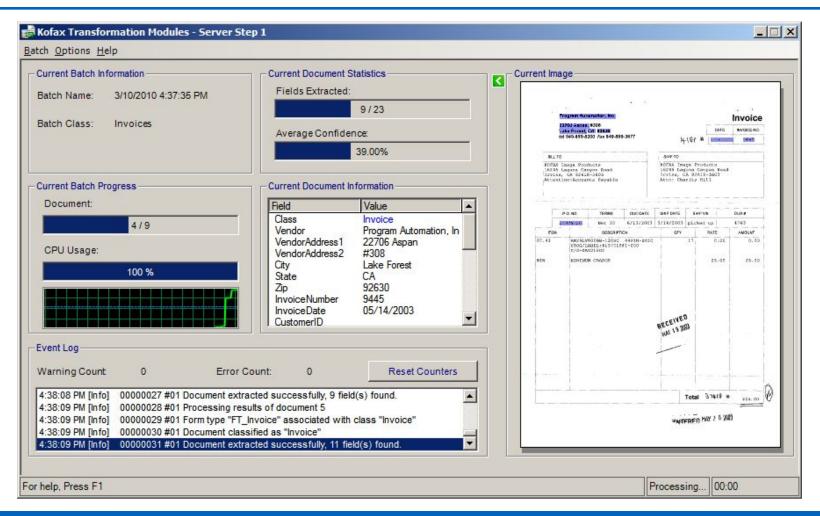


#### **Project Planner (Requires Professional License)**

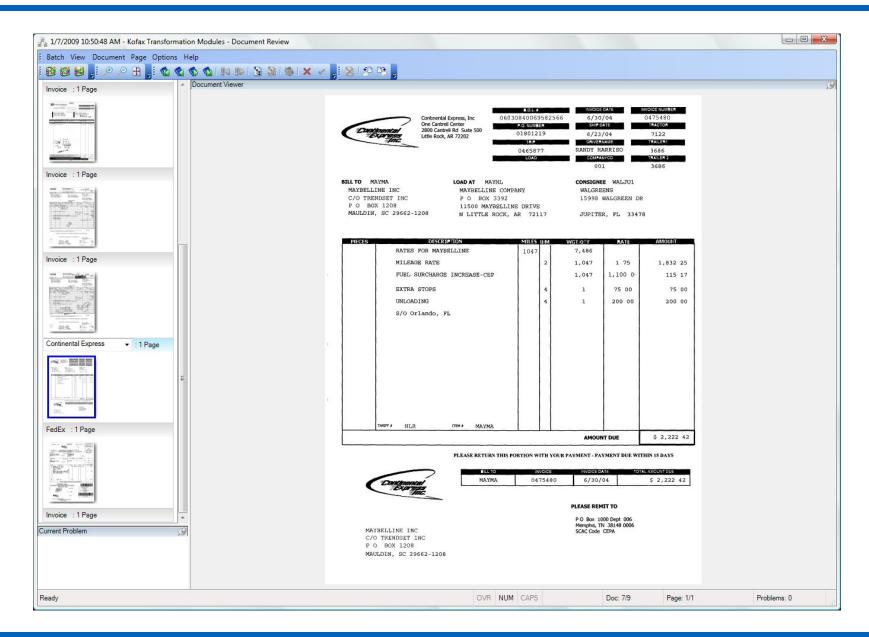


#### **KTM Server**

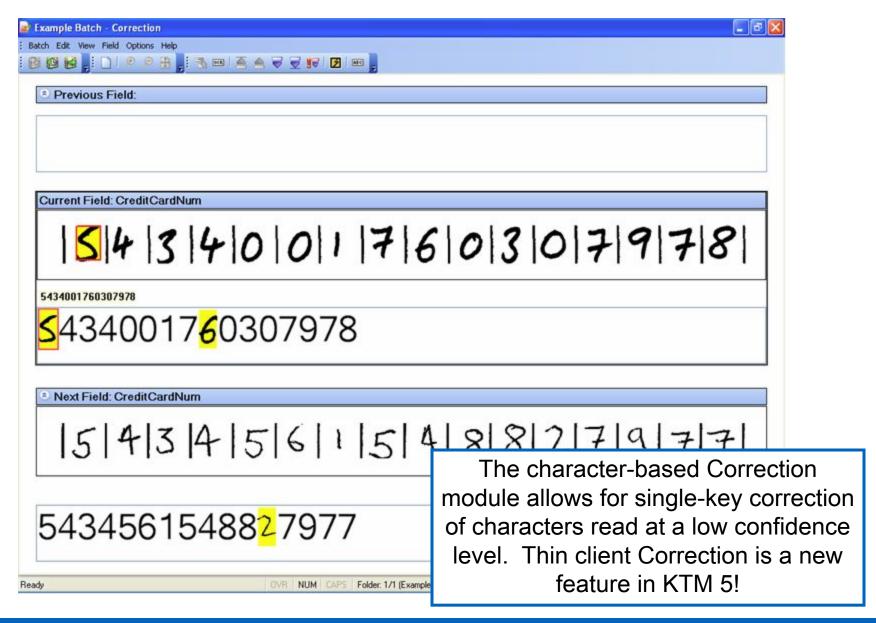
The Server may run as an application or as a self-starting Windows service. Up to 16 services may be used. It classifies, separates and extracts. With the Pro Upgrade, you may also want to run separate instances of the Server: one for classification and separation and one for extraction.



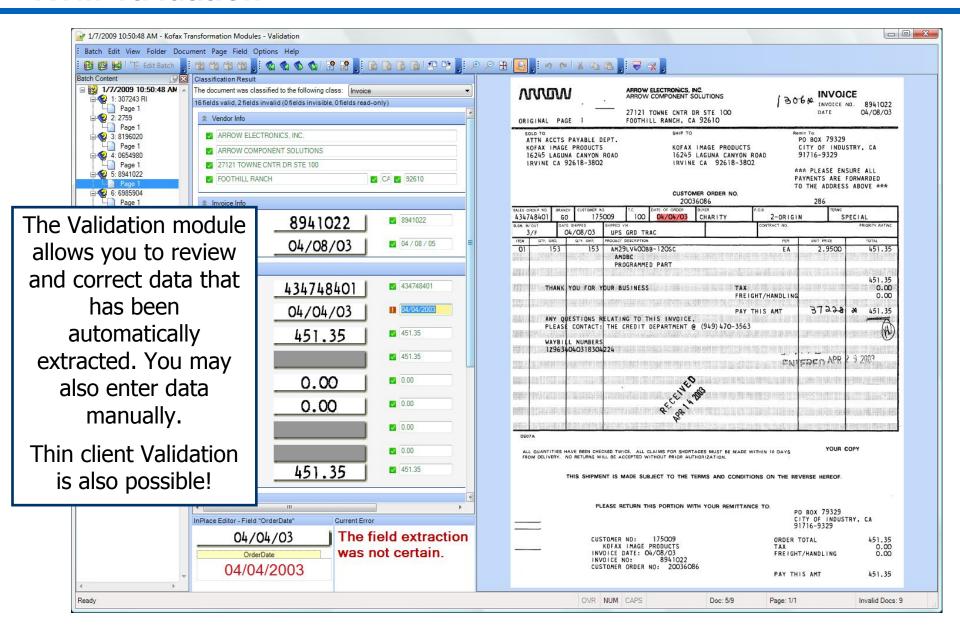
#### **KTM Document Review**



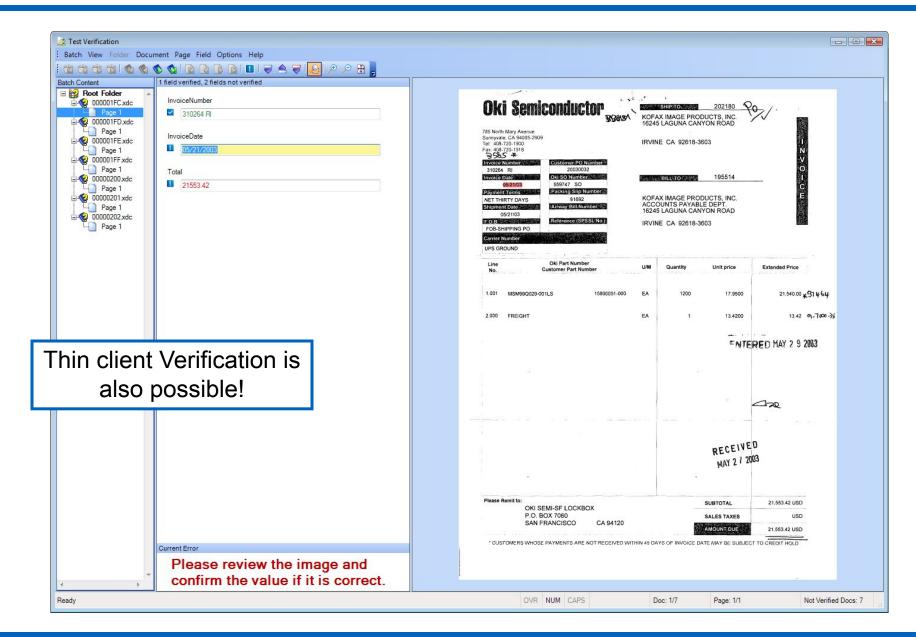
#### **KTM Correction**



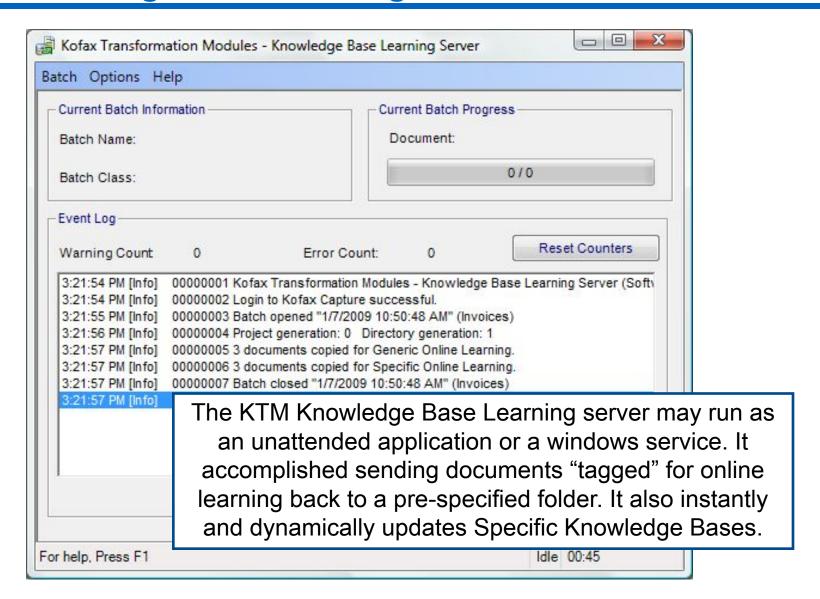
#### **KTM Validation**



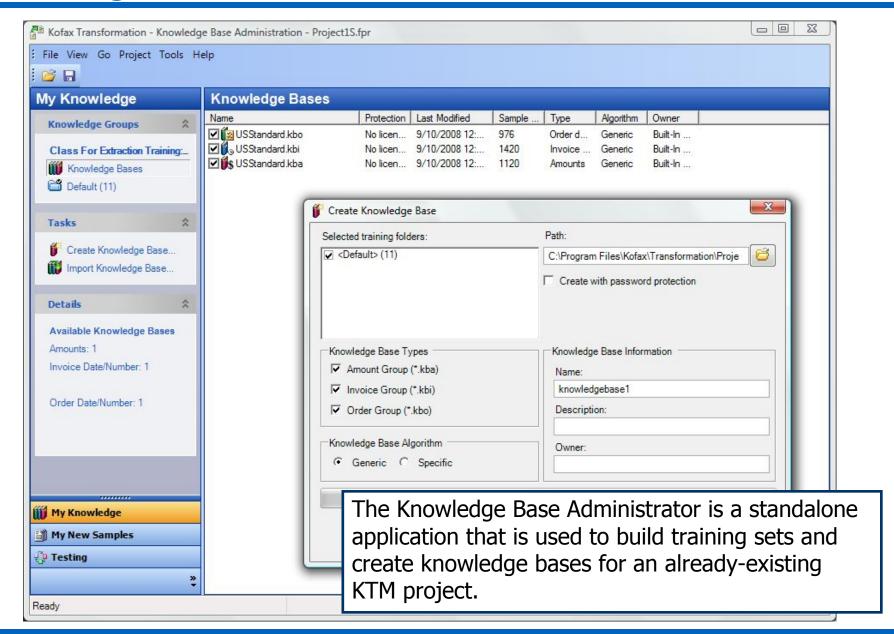
#### **KTM Verification**



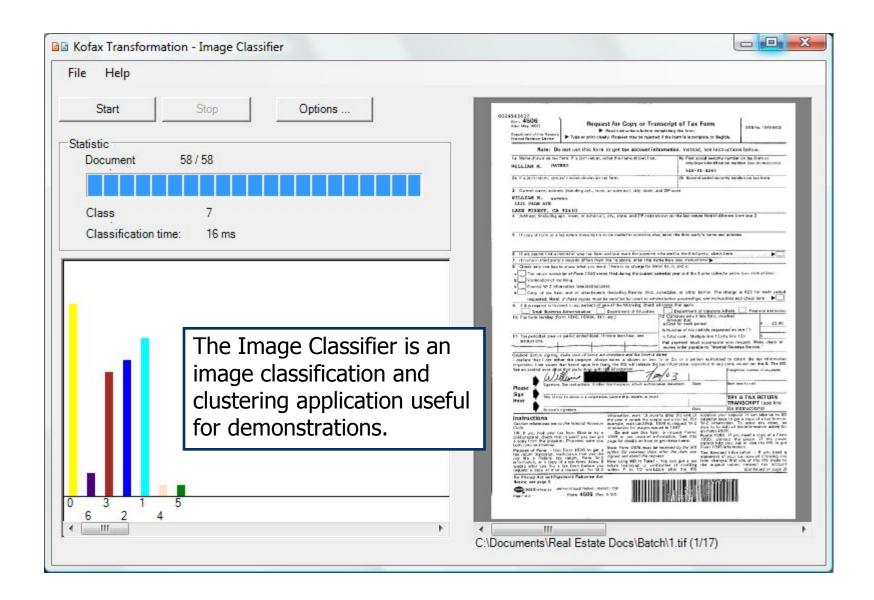
#### KTM Knowledge Base Learning Server



#### **Knowledge Base Administrator**

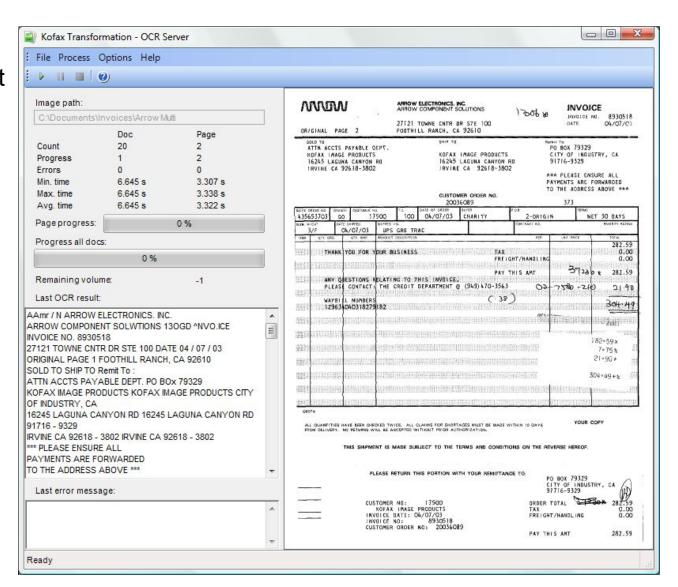


#### **Image Classifier**



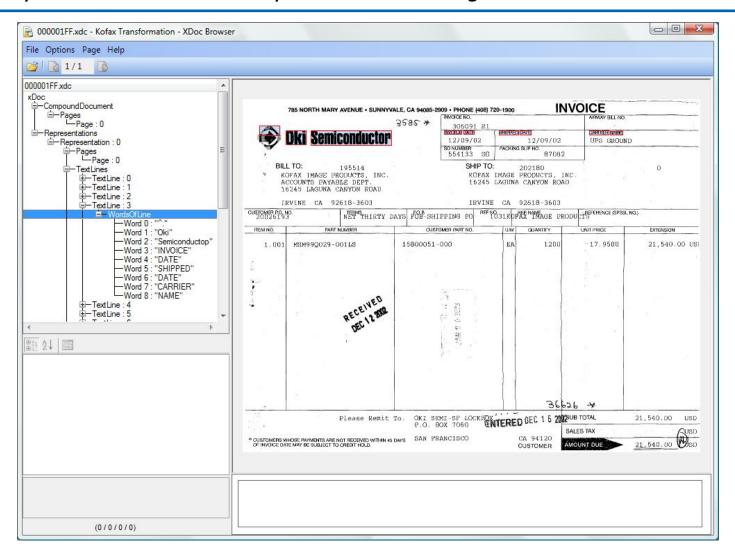
#### **OCR Server**

The OCR Server allows you to create OCR output (including Xdocuments) outside of Project Builder and without having to process batches through the Kofax Capture workflow.

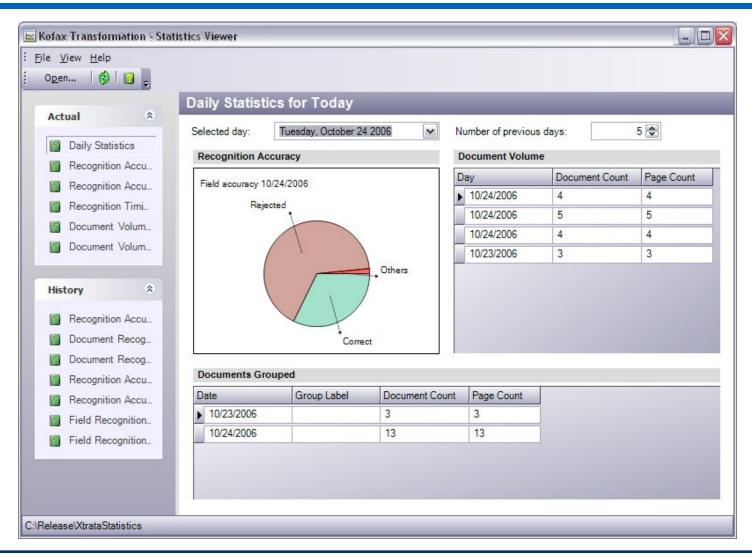


#### **XDoc Browser**

The XDoc Browser provides direct access to the results of automatic processing on a document-by-document basis. It may be used for testing and demonstration.

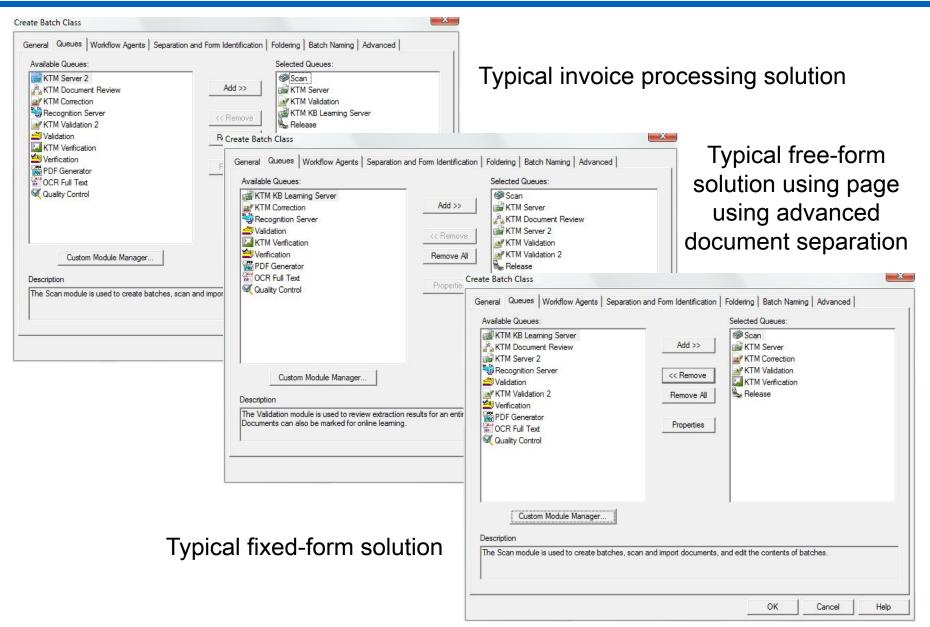


#### **Statistics Viewer**

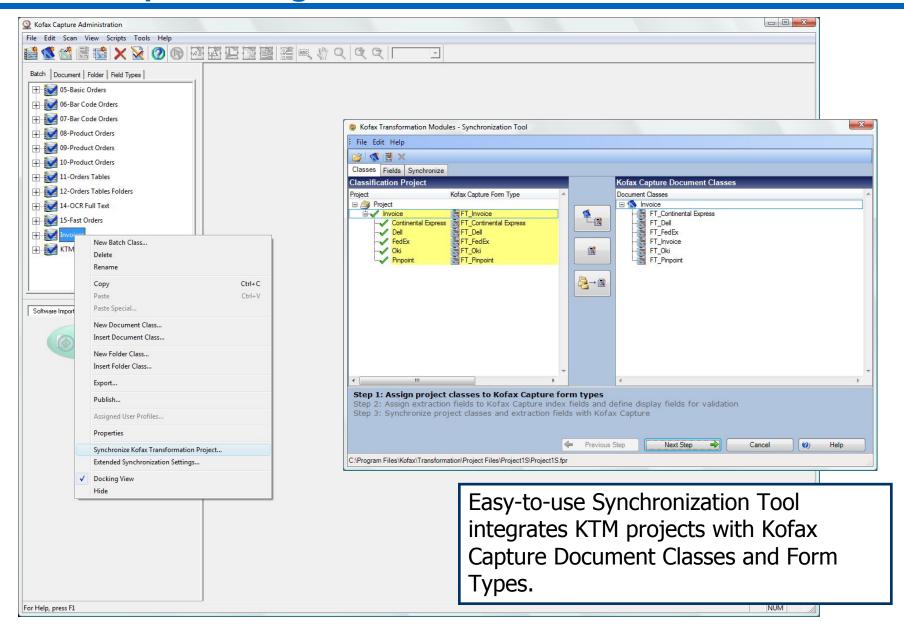


The Statistics Viewer allows you to view summary statistics of the results of automatic processing as you set up and modify your projects.

#### **Kofax Capture Batch Workflow**



#### **Kofax Capture Integration**



## KOFAX Transformation Modules 5

in ACTION



