

SLASHING MANUAL INPUT WITH ORACLE DOCUMENT CAPTURE

Dmitri Khanine, StellentExperts.com

ABSTRACT

If you ever wondered what kind of business impact you can achieve – by relieving users from having to fill in all these fields on the check-in form that make them cringe, this session (and this whitepaper) may literally bring home hundreds of thousands of dollars ... and get your business people dancing with joy! Attend this session for a live demonstration of how using Oracle Capture can virtually eliminate dreaded manual input, skyrocket the quality of your content and explode user productivity.

INTRODUCTION

Hi, I'm Dmitri Khanine from ECM Solutions, elite special forces unit focusing almost exclusively in Oracle UCM space. I'm the author of The Oracle UCM Handbook and I also run the Independent Oracle UCM Knowledge Center. We've just completed our second large Oracle Capture rollout in the last 12 months and I'm here to share the tips, tricks, insider information and time-saving "gothchas".

Let's start by quickly looking at why so many Content Management implementations fail... and also...

WHAT SALES PEOPLE ARE GETTING ... RIGHT?

You've probably heard that sales pitch about controlling the scary bloat of ever expanding mass of unstructured content – paper, email, file shares and how important it is to help your business people to really access the information they need, when they need it – from anywhere. And you know what – that part is right. I agree with the sales guys. Unstructured content really is hard and expensive to manage. I've personally seen clients losing hundreds of thousands of dollars having micro-chips manufactured to an outdated specification, 'cause the production team looked in a wrong folder in an unwieldy SharePoint repository that gone out of control.

So there must be a simple solution, right?

FAST RELIEF TO INFORMATION BLOAT PROBLEM... OR NOT...

Here's where many good intentions fall flat on a cold hard concrete of reality... the tool alone will not solve your business problem! No matter what tool you pick.

People install Oracle UCM, but unless they capture the metadata – the system is not much better than a plain old file system. Yes, they realize that and they define the metadata they would like to capture, so their check-in screens end up something like this screenshot on the previous page.

Problem solved, Right?

Uh... not exactly.

You see, when a contributor sees a form like that, they get terrified... or annoyed... or both. And they try to get away with as little input as possible. So they skip as

many fields as they can and when the system pops up a “value required” thing at them – they enter garbage.

So is it really practical to collect more than just a handful of metadata fields on check in?

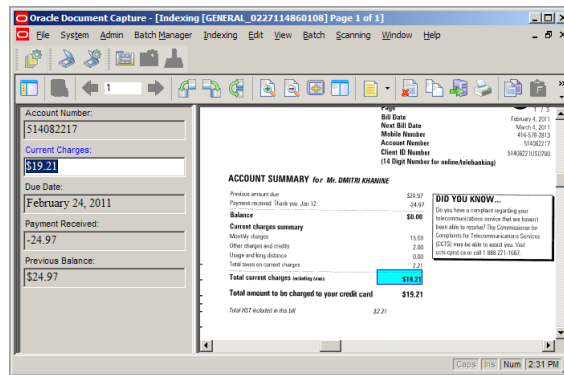
Probably not, if you do it by hand... But who said you ought to do it by hand?

INTRODUCING ORACLE CAPTURE

Let’s say, we’re checking in a monthly invoice and have five fields we’d like to capture. How would you like to – instead of putting them on the check in form – let your contributor to look at the actual document, side-by-side with a check in form... and have OCR to take its best guess at pulling that data and populating the form – before the contributor even gets to look at it?

Well, this is exactly what Oracle Capture is designed to do – pretty much out of the box.

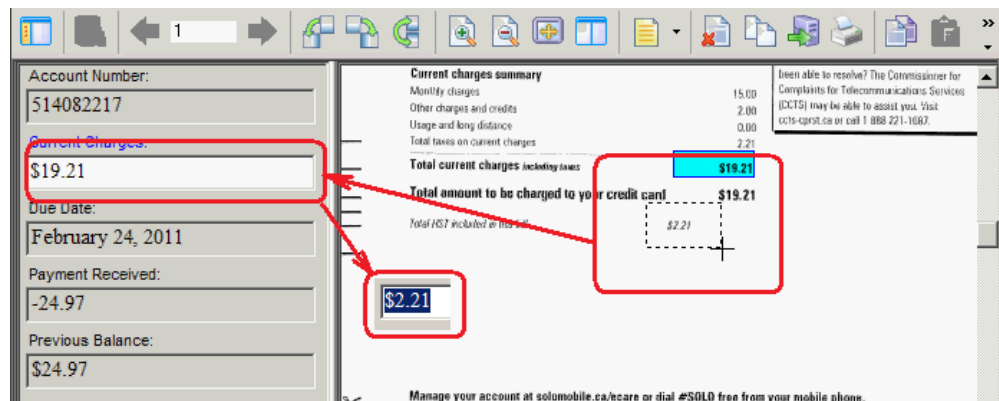
Take a look at the screenshot below:



Not only contributor gets to see the document and the metadata fields side-by-side, each field has a corresponding OCR zone defined – and Capture is actually extracting the data out of this zone and puts it into the field.

When she moves to the next field, that field’s zone is highlighted in blue and she gets to see where the value in the field came from.

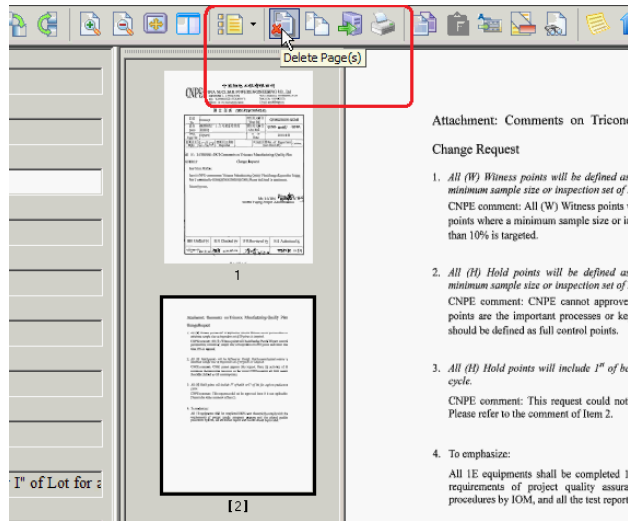
If the document was skewed or poorly scanned and the zone misses the value, contributor can simply drag and drop the value (See screenshot below)



Just select the area containing the text that you want to see in the field with your right mouse button and see the OCR extract the text out of the area and place in the field for you! How many fields do you think you can capture now?

You can also combine multiple documents into one and you're free to remove junk pages as you see fit. Imagine what would it take to delete a page from a PDF document during normal Content Server check in. It's simply not practical so these junk pages end up in repository...wasting time when searching for information... and folks print them every time and waste trees.

All solved in Oracle Capture at a click of a button (See screenshot below)



I hope you're getting a feel for what it can do. So now it would be a perfect time for some...

REAL LIFE STORIES

I'm going to show you a couple of real life examples, so I want to quickly prove to you that this tool actually works.

SLASHED HALF OF THE LOAD OF OVERWORKED DOCUMENT CONTROL PEOPLE

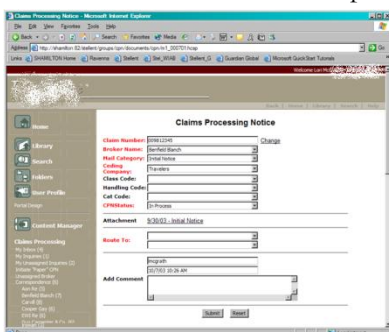
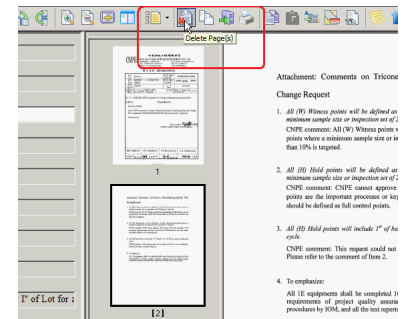
One of the projects we recently completed was automating workflow and document control for a nuclear power station operation management provider. As part of doing business, they exchange a ton of technical specifications and formal transmittals with their government clients and sub-suppliers. Prior to implementation, information was duplicated, not easy to organize and retrieve and they didn't have a structured workflow process.

Before using Capture, average time spent per document was benchmarked at 2 min 17 sec – on a 15 field check in form. With using Capture for extracting verifying and correcting pre-filled values – time spent per document dropped to just over 33 seconds!

That's over two hours of added productive time (out of estimated 4) per Document Control person per day – or a *killer 200% increase in productivity!*

BUMPED UP PRODUCTIVITY BY 75%

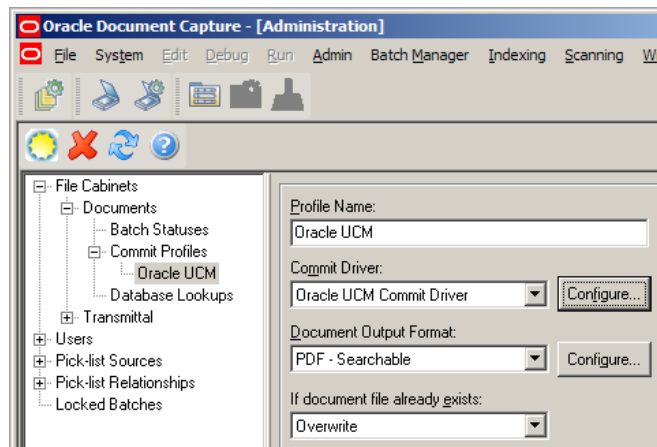
Another our recent project I'd like to mention is the one where we've automated document storage, acquisition and workflow for a mid-sized insurance company. They have significant volume of incoming claim-supporting correspondence – mostly scanned documents coming in by email.



Over the years, they've tried more than once to extract meaningful metadata fields such as claim number and policy id – on check in – with various degrees of success.

But if your documents are coming in as MS Word and PDF on a shared drive or email attachments - you don't have to print and re-scan them. Simply use the Import Server that will gladly import anything that can be virtually printed. You can also tell it to watch a folder, and drag and drop documents there – and see them appear in Oracle Capture.

Ok, but what happens when the user clicks “Commit” button on her Capture toolbar? There're many options, but the one I want to focus on is having the document *committed to Oracle UCM* (See screenshot below)



It's easy enough to configure, and you can have a searchable PDF/A committed to the Content Server... in addition to being able to map your Capture metadata fields to Content Server.

And now here's a feature so profound, I had to put it into its own section:

RAPID CUSTOMIZATION FACILITY

If you need to apply business rules on Content Server check-in - you can use profiles and rules, but they are quite limited. Custom Java components take a long time to write, plus you risk having to re-write them after every version upgrade...

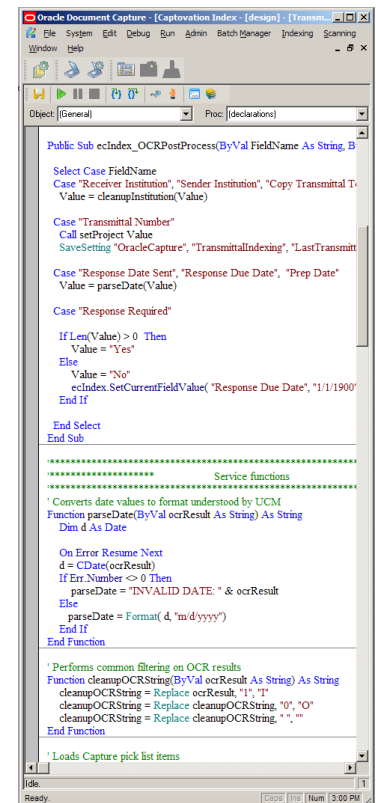
Good news! Capture offers you quick, powerful and easy to use alternative – *Basic Macros*. Just like in MS Word, Excel and Outlook, you can verify and alter data in the fields, handle events like page load (and Capture's own commit and OCR-related events), populate dropdown values and display custom dialogs.

Another thing you can do is go back to the database (a Content Server one if you like) and bring back a bunch of metadata values based on just one field. If you do that – all you need to grab off the scanned image is just one field! Claim number brings back the Broker Name, Adjuster, Policy Number and so on.

And once we started to talk about writing macros in Oracle Capture – there's something else I have to mention...

DON'T SAY I HAVEN'T TOLD YOU!

The first time we used Capture a few years ago – we quickly flipped through developer manuals and it only took us 15 minutes to install the tool itself, so went for a beer to celebrate! What an easy tool! Phew! We felt like we've almost completed the project.



Same time next week.... Bitter disappointment! Nothing works! Can't even create an object! Assigning values to properties.... no result! Enumerating collections.... collection is empty! EMPTY? But I'm looking at the damn batch! It has 10 pages! How can the pages collection be empty?

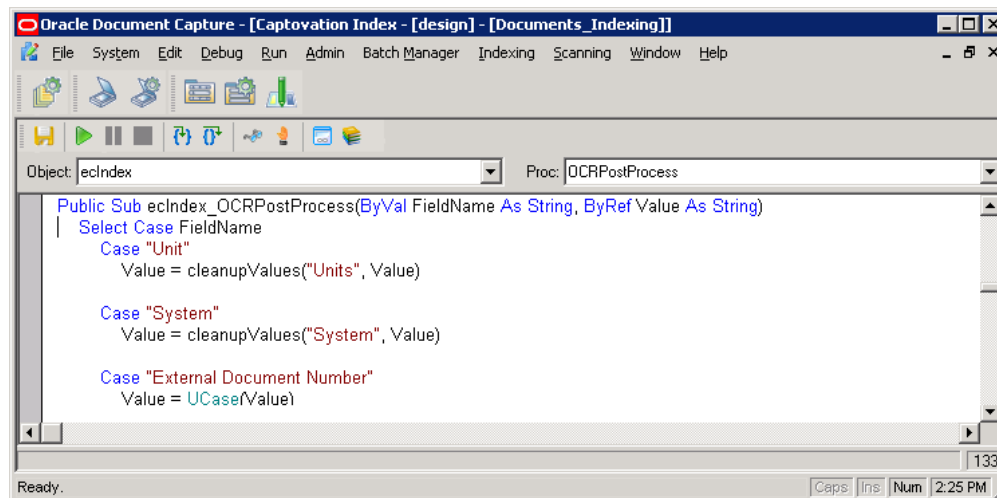
Pounding on keyboards... Screaming... *A few days* of fun!

But you know what, it all started to work after we've learned a few tricks. I'm actually going to give you those in the next section. Macros do seem really simple at first, and they are, they do take a little time and experimentation to get used to and get them to work. So here're the few tips I wish we had in front of us back then.

QUICK POINTERS FOR FRUSTRATION-FREE MACRO CODING

This is a list you want to study before you write your very first macro in Oracle Capture:

- Start with *ecIndex* > *OCRPostProcess* event. This is the most useful event handler and one of the easiest ones to use (See screenshot below). Simply check whether this is the field you need to update and change the *Value* variable. You can also use the value that came back from OCR as a key for a database lookup – and replace the field's value with the result.



- When reading documentation, you'll see a lot of references to *Connection* object but never an example of how to obtain a reference. It's actually easier than it seems. Use *eCaptureMacro* > *Connect* event handler and store the reference to *Connection* object in a variable for future reference:

```
Public Sub eCaptureMacro_Connect(ByVal Connection As Object)
    Set conn = Connection
End Sub
```

- Check out the signatures for other event handlers, such as *Batch Open*, *Batch Pre-Commit* and *Field Lost Focus*. They give you other important object references that you might need to use later in your code... and help you apply business logic after the user has hit "Enter" or used her mouse to go on to the next field.
- Right after obtaining reference to a collection like pages, indexes, file cabinets and so on, be sure to call a Refresh method before attempting to enumerate the collection. If you don't – collection will have no records in it! So remember to call Refresh, just like I did in the examples below:

```
myBatch.BatchPages.Refresh
```

```
cab.IndexDefinitions.Refresh
```

```
pickList.PickListItems.Refresh
```

and so on...

Here's a complete snippet that sets *cab* to a reference to a file cabinet named "Documents":

```
conn.FileCabinets.Refresh
Dim cab As Object
  For Each cab In conn.FileCabinets
    If cab.Name = "Documents" Then
      Exit For
    End If
  Next
```

- When modifying the index values on multiple pages, remember to call *page.Persist* method. If you don't the changes will not be committed.
- When all you need to do is simply get or set the value of a metadata (index) field on the current page – simply use these calls:

```
ecIndex.SetCurrentFieldValue("Document Name", myDocName) – to set the value of “Document Name” field to the content of myDocName variable ...
```

or

```
if ecIndex.GetCurrentFieldValue("Apply to all pages") = "Yes" Then ....to check if the value of “Apply to all pages” field is set to “Yes”
```

That's it! Believe it or not, I've probably saved you at least a day worth of trial and error. You're now ready for frustration-free macro coding experience.

And now that we've looked at these and the core features of Oracle Capture, let's see what it takes to really crash it on results on the corporate level. So here're the...

3 KEYS TO ACHIEVING MAXIMUM BENEFITS WITH ORACLE CAPTURE

This section will give you some important pointers on bridging the gap between a really helpful tool and massive enterprise-wide business results. I'm giving you 3 critical keys to a successful corporate rollout of Oracle Capture. Ignore them at your own risk.

KEY #1 - IT WON'T WORK IN ISOLATION

If one of the departments starts using Capture and correctly populate a bunch of metadata fields... and other departments continue to grudgingly punch in required values by hand – the good and the bad entries will all end up in the same repository and you won't be able to tell the good from the bad.

If you want to see some real results – you may be looking at an enterprise-wide rollout and, possibly, some data cleanup work down the road.

KEY #2 - THE MORE YOU USE – THE MORE YOU GAIN

Yes, it does sound common sense, but it's true. The more of your content you have in electronic form, available for searching – the greater the power you give your business people. The more of the relevant metadata fields you define – the easier it is to search, manage and orchestrate your processes.

Once all of your content is in – you can also control its retention. Delete or archive what’s not being used. Eliminate multiple versions of the same document....

This is why companies undertake back file conversions, scan and eliminate paper copies, get rid of their mail rooms and off-site storage. If it takes a day or two to obtain a paper copy – what’s the benefit of having that copy in a first place? Sometimes, spending even 5 minutes searching for it will be posing a risk of making a wrong business decision or missing out on important opportunity.

KEY #3 - ONE SIZE... WON'T FIT ALL

Here’s the last one, but by far not the least. To really get what the Capture has to offer - the true benefits, like the smooth, tailored user experience that takes you to the next level of productivity – you got to customize the product ... Add business logic... clean up your OCR data... make your metadata fields “smart” – to handle partial matches, pop up custom dialogs, link the fields, capitalize on database lookups... An extra day of development may well bring you hours worth of saved contributor effort – day in and day out.

BEYOND THE TURBO-CHARGED CHECK-IN

That’s great, but what if you don’t want someone to sit there and verify every document? What if you’re looking at a back file conversion of thousands of documents – and you want to have it run unattended?

You really have two options – use bar codes and Oracle Capture Recognition Server or step up to Oracle Forms Recognition

If you can afford to use database lookup and macro code to infer the values of your metadata fields from one or two key fields, the Recognition Server that comes with Oracle Capture will be an ideal solution. Simply add a bar code to your document – to represent the key value, such as your policy or file number - and have a database lookup pull out the rest. That gives you fully automated forms recognition. No human intervention required.

If you need to process a variety of forms and you want the system to decide what form it is – and pull out the correct metadata values – you’re looking at a totally different animal and you got to use a different tool. This tool called Oracle Forms Recognition or OFR.

While Capture is based on a set of user-selected document profiles, each with a fixed OCR zones, OFR comes with intelligent full-form OCR and automatic matching to a document profile. That said, its more expensive and more work to setup and “train”. So once again, one size does not fit all. Pick your tools to match your project.

CONCLUSION

You’ve just learnt about a killer tool that can quickly put controls around a growing mass of your corporation’s unstructured content and take your business people to the next level of productivity. You’ve seen how Oracle Document Capture slashes manual input with its simple, intuitive and flexible interface... and you’ve seen the actual numbers that come with the massive time savings and productivity gains this tool produces in real life settings.

I’ve shown you what tools you have to customize Oracle Capture to fit to your own requirements and I’ve also given you some insider tips and things you need to be aware of when writing Basic code in Oracle Capture.

You’ve also seen the three keys to achieving the maximum benefits with Oracle Capture and what options do you have when going beyond a “turbo-charged check in screen” solution.

For more information visit ECM Solutions web site at www.stellentexperts.com

REFERENCES

1. Oracle Content Management web site: <http://www.oracle.com/technetwork/middleware/content-management/overview/index.html>
2. The Oracle UCM Handbook (Packt Publishing, 2010): <https://www.packtpub.com/oracle-universal-content-management-ucm-handbook/book>

3. Oracle Document Capture Administrator's and Developer's guides - http://download.oracle.com/docs/cd/E10316_01/capture.htm