Logo will appear here

Microfiche scanning and indexing

Description | Mockup | Demos | Project Artifacts

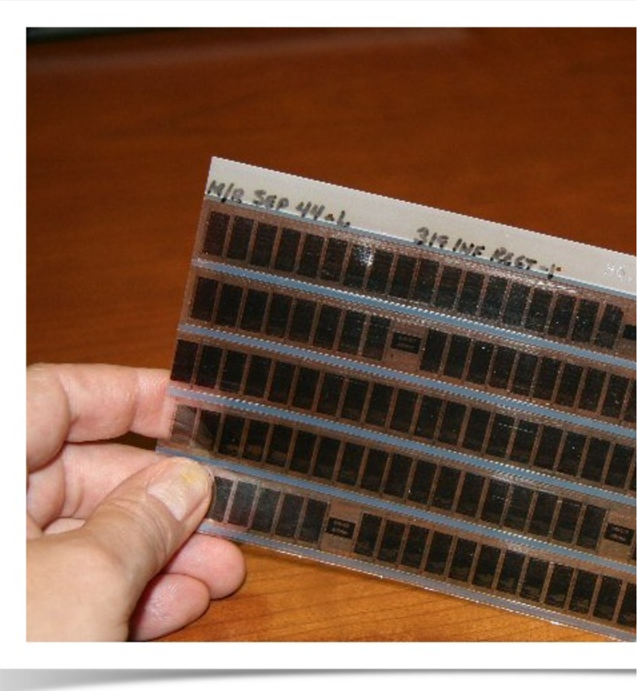
Description:

Explore technology options and potential investment required to provide Microfiche conversion capabilities for converting to digital images for clients.   There is so much Microfiche in the world that requires scanning you could literally, "Do this all day long" - Jim Walsh - DST.

Background:

Many of our clients have tons upon tons of old Microfiche sleeves that they would pay to have digitally imaged and stored in EFS or a similar document store.   These would also require indexing for lookup purposes, this indexing could be performed by a human workforce with onshore or offshore, or via computer automation if appropriate.   Indexing would be something along the lines of account number or SSN.

Microfiche are postcard sized transparent cards that contain multiple shrunken images that when used to a reader device are enlarged by projection and displayed onto a large screen.



Objective:

Determine if we are able to cost-effectively provide a product offering based on hardware/software available to provide Microfiche to digital imaging with data indexing.

Plan:

Requirements

1. Digitize microfilm/fiche:
   a. DST (working with Jane Cory in our Microfilm room in CT)
   b. BFDS (as of 10 May 2018, BFDS is to ship a bunch of films to be stored in PD)
   c. AXA (40 million images - 8.5 million fiche & 5,500 rolls of film)

2. Build UI (Angular / SpringBoot cloud-based project)
   a. The first phase is to simply digitize images and provide a UI to enable users to scan through images very much like using a microfilm/fiche reader machine
   b. Subsequent phases will include abilities to search on metadata extracted by the batch processing of the images

3. Batch Processing - a mix of Computer Vision, AI and Brute Force methods to:
   a. OCR typed text
   b. extract handwriting
   c. extract meta-data such as account numbers, SSNs, names
   d. detect "batch" separations
   e. and more

Scanning
Using 3rd Party, American Micro , to do the scanning. At a minimum, we need them to scan each real or card into its own folder and associate that with its label.

Indexing for retrieval options

Day 1 - Setup MySQL between UI and storage to provide mappings to digital images. Mappings mimic info. used to find films manually today, and varies by client. For example, document type, year, doc id ranges, etc. The mappings are essentially the digital equivalent to the film/fiche labels. These mappings are used to populate the filtering dropdowns in the UI.
Day 2 and beyond - Expand MySQL to capture results of the batch processes, and enhance the UI to do searches against those results

Storage
Storage is to be in the EDP for ¢ rather than on file systems for $.

Initial Plan / Research from 2015...

Methods:

You could foresee three components to a product offering

Hardware to scan the fiche to digital - potentially with the ability to index at the time of scan? Research options to buy/build.
Software to potentially automate the indexing of the images - ambitious.
Software consisting of a user interface that displays the digital images of the fiche and allows a human operation, potentially offshore to enter the indexing information with the keyboard/touchscreen/speech etc.

Results:

Initial Market Research of the marketplace performed by Zhejiang University/HengTian - **complete**
Scanning experimentation with flatbed - awaiting samples - **not feasible,** special scanners required.

Requirement
1. Turns out the client is AXA and has mostly Microfilm with some Mircofiche - 40 million images (8.5 million fiche & 5,500 rolls of film)

Scanning options
2. Met with Mail center and they are open to housing equipment and staffing, they may have some spare staff capacity. A scanner for this is around 40k. An example of us doing the film scanning with our own hardware purchased from NextScan. @300DPI, 8-hour shifts… 5,500 rolls would take 7.5 months on the Eclipse 400 model, and 5.0 months on the Eclipse 600 model.

3. Another option is to outsource to somebody like Scan America (in Lawrence, KS) they charge around 1 cent per image - a little more if they do the indexing. Luke visited their site last week. So, 40k to scan offsite. Scan America suggested 4-5 weeks as they have 5 scanners and can rent more if needed.

Indexing for retrieval options
4. We could do a hybrid using ScanAmerican and the KC mail room and touchscreen software to index - the software supports indexing.
5. Secure Task may be an alternate option for indexing. Pricing information collected. Indexing requirement is mostly handwriting making OCR a non-starter.
6. We build something quickly, new out of R&D
7. We have Scan American do the indexing

Storage
6. Storage-wise, the client seems to favor AWD. This would also match our current internal business processes for Fiche. Haven't gotten costs yet. Maybe a tie into Devin's storage initiative.

Current Business Process

* Current process at DST and AXA is to scan into AWD using a Film/Fiche reader by printing out and then sending to the mail center on client request of an image. Print out is sent via the postal service and image is in AWD if it needs to be sent again. We will end up creating a solution to migrate the Fiche/Film DST store for clients as part of this.

Offshore
* HengTian did some free research and found nothing of real interest in the China market. We contacted some India providers and scared them off with the volume I think, or they just aren't serious.

Next Steps
* We are waiting for samples from AXA - requested last week.
* At some point this wouldn't be an R&D type of project, but one we could either push through and complete as a one-off looking to do something novel along the way for patient/client stickiness or work with another group to make this a repeatable process for other clients to buy.

Progress:

Initial research was presented to mid-November 2015.
This has been a paused project for some time due to resource constraints but has been picked up during 2017 summer internships within R&D and also with two days per week experienced MIC interns.
Two attempts at the creation of automated numeric digit indexing have been performed. One in Hyderabad as part of the Above & Beyond R&D program, and a second out of the Toronto R&D lab. Both were using AI neural networks. Both have proved it possible to achieve automated indexing at comparable levels of human beings but at greatly reduced costs.